INVESTIGATION · METHOD AND THE FUTURE

Measuring Investigative Effectiveness Beyond Case Clearance Rates

Clearance rates, the share of cases closed or solved, are the most common measure of investigative performance, but they say little about accuracy, fairness, harm prevented or what was learned. This article explains the limits of clearance rates, how targets can distort behaviour, and a broader set of measures for judging whether investigations are effective.

ModQu research article · 1,970 words · 20 references (APA 7th edition)

How do we know whether investigations are working? The most common answer, in policing and beyond, is the clearance rate: the proportion of reported cases that are solved, closed or result in a charge. Clearance rates are easy to count and compare. But they measure only one thing, whether a case was closed, and they can be closed in ways that have little to do with finding the truth. An investigation unit could achieve a high clearance rate by focusing on easy cases, by closing difficult ones quickly, or even by charging the wrong person. Meanwhile, the qualities that matter most, such as accuracy, fairness, the prevention of harm and the confidence of the public, go unmeasured. This article explains the limits of clearance rates, the ways in which targets can distort investigative work, and a broader approach to measuring effectiveness.

What clearance rates measure

A clearance rate counts cases that reach a defined outcome, such as an arrest, a charge or an administrative closure, as a proportion of cases recorded. It is useful for tracking workload and broad trends. But research has long shown that clearance depends heavily on factors outside investigators' control. The RAND study of criminal investigation found that whether a case was solved depended mainly on the information available at the outset from victims and witnesses, rather than on the effort or skill of investigators (Greenwood et al., 1977). Eck (1983) found that investigative effort mattered for a subset of cases with moderate initial information, but that many cases were either solved immediately or not solvable at all. Wellford and Cronin (1999) identified specific investigative practices associated with homicide clearance, showing that investigative work does matter, but the overall rate still reflects the mix of cases as much as investigative quality.

Clearance rates are therefore a poor comparison between units or over time unless the type and difficulty of cases are taken into account. A unit that handles many cases with named suspects will clear more than one that handles many cases with no leads, regardless of skill.

How targets distort behaviour

When a single measure becomes a target, people change their behaviour to meet it, sometimes in ways that defeat its purpose. Campbell (1979) observed that the more any quantitative social indicator is used for decision-making, the more subject it becomes to corruption pressures and the more apt it is to distort the processes it is meant to monitor. Kerr (1975) described the "folly of rewarding A, while hoping for B": organisations often reward measurable behaviour while hoping for something else. Muller (2018) documents how reliance on metrics in policing, health, education and business has led to gaming, short-termism and neglect of what is not measured. Bevan and Hood (2006), studying targets in the English health service, found evidence of gaming, including meeting targets on paper while leaving underlying performance unchanged.

Policing provides clear examples. Eterno and Silverman (2012) reported evidence from surveys of retired officers in New York that pressure to show falling crime under a performance management system led to the downgrading of crime reports. In England and Wales, a parliamentary committee found that police recorded crime statistics had been distorted by pressure to meet targets, including the under-recording of crime and the misclassification of offences, and that the statistics could not be relied on (Public Administration Select Committee, 2014). Moore and Braga (2003) argued that performance management systems in policing can improve focus and accountability but must measure a wider range of values than crime numbers if they are not to distort practice.

For investigations, a clearance target can encourage investigators to prioritise easy cases, close difficult ones prematurely, or press for confessions and charges when the evidence is weak. It can also discourage the honest recording of cases as inconclusive.

What effective investigation means

A broader approach starts by asking what investigations are for. Most aim to establish what happened, hold the right people accountable, protect victims and the public, treat everyone fairly, and do so efficiently and in a way that maintains public confidence. Each of these aims suggests its own measures.

Dimensions of investigative effectiveness

A balanced set of measures reduces the distortions caused by any single target

	What it asks	Example measures
Accuracy	Were findings correct?	Findings overturned on appeal or review; later exonerations
Harm reduction	Was harm stopped or prevented?	Harm-weighted outcomes; repeat victimisation; losses recovered
Fairness	Were people treated fairly?	Complaints upheld; disclosure failures; time under suspicion
Timeliness	Was the work done promptly?	Time to key decisions; backlog age
Quality of process	Was the work done properly?	Peer review scores; file audit results
Confidence	Do people trust the process?	Victim and witness satisfaction; willingness to report
Learning	Did the organisation improve?	Lessons implemented; recurrence of the same failures

ModQu analysis based on Moore & Braga (2003), Sherman (2013) and Kaplan & Norton (1992) ModQu.com

Figure 1. Effective investigation has several dimensions. Measuring them together gives a fuller picture than clearance rates and makes it harder to meet one target at the expense of the others. Diagram by ModQu.

Accuracy. The most important quality of an investigation is that its findings are correct, but accuracy is hard to measure because the truth is often unknown. Indirect measures include the proportion of findings overturned on appeal, review or reinvestigation, and later exonerations. Gross et al. (2014) estimated that at least 4 per cent of people sentenced to death in the United States were likely innocent, showing that error rates can be substantial even in the most carefully reviewed cases. Garrett (2011) used exonerations to identify the investigative practices most associated with wrongful convictions.

Harm reduction. Investigations should be judged partly by the harm they prevent or stop. Sherman (2013) argued that policing should target, test and track its work against evidence of harm, and Sherman et al. (2016) proposed weighting crimes by their seriousness, using sentencing guidelines, so that solving one serious case counts for more than solving many minor ones. Other measures include repeat victimisation, losses recovered and the disruption of networks. Braga et al. (2011) argued that investigative units should be judged by their contribution to controlling crime, not only by individual case outcomes. Sparrow (2000) argued that regulators should measure their success by the problems they solve, not by the number of enforcement actions.

Fairness. Fair investigation means treating suspects, victims and witnesses with respect, following due process and disclosing relevant material. Measures include upheld complaints, disclosure failures, the length of time people remain under investigation, and differences in outcomes between groups.

Timeliness. Delays harm everyone: evidence degrades, memories fade, suspects remain under a cloud and victims wait for answers. The age of open cases and the time taken to reach key decisions are useful measures.

Quality of process. Peer review and file audits can assess whether investigations followed sound practice: whether evidence was preserved, alternative hypotheses considered, decisions recorded and findings supported.

Confidence. Investigations depend on public cooperation. Sunshine and Tyler (2003) and Jackson et al. (2013) found that people's trust in the police and willingness to cooperate depend strongly on whether they are treated fairly. Surveys of victims, witnesses and the public measure this directly, and trends in reporting may reflect confidence.

Learning. An effective investigative organisation improves over time. Measures include the proportion of review recommendations implemented and whether the same failures recur.

Measuring accuracy in practice

Because the true answer in most cases is never independently known, organisations need practical ways to estimate accuracy. Several are available. A sample of closed cases can be re-examined by experienced reviewers who were not involved, who assess whether the findings were supported by the evidence and whether plausible alternatives were excluded. Findings that are challenged, through appeals, tribunals, complaints or later information, can be tracked to see how often they are upheld. Cases in which new evidence later emerges, such as a cold case resolved by DNA, provide occasional direct checks. Where the same question is investigated by two bodies, for example a regulator and a court, agreement between them is informative. None of these measures is perfect, and each covers only some cases, but together they give an organisation a far better sense of its error rate than clearance figures, which say nothing about whether the right answer was reached.

Using measures to improve, not to blame

How measures are used matters as much as which measures are chosen. When numbers are used mainly to rank teams and punish those at the bottom, people learn to manage the numbers rather than the work, which is the pattern Campbell (1979) and Bevan and Hood (2006) described. When measures are used to identify problems, prompt questions and test whether changes work, they support learning. Practical steps include sharing results openly with the people whose work is measured, discussing the reasons behind changes in the figures rather than reacting to the figures alone, combining numbers with case reviews, and protecting staff who report errors or gaming. Measures should be reviewed with frontline investigators, who often know best whether a measure reflects good work or merely rewards activity that looks good.

Building a balanced set of measures

Kaplan and Norton (1992) proposed the balanced scorecard as a way for organisations to track several dimensions of performance together, so that improvement in one is not achieved at the cost of another. The same principle applies to investigation. A balanced set might include a small number of measures from each dimension above, reviewed together.

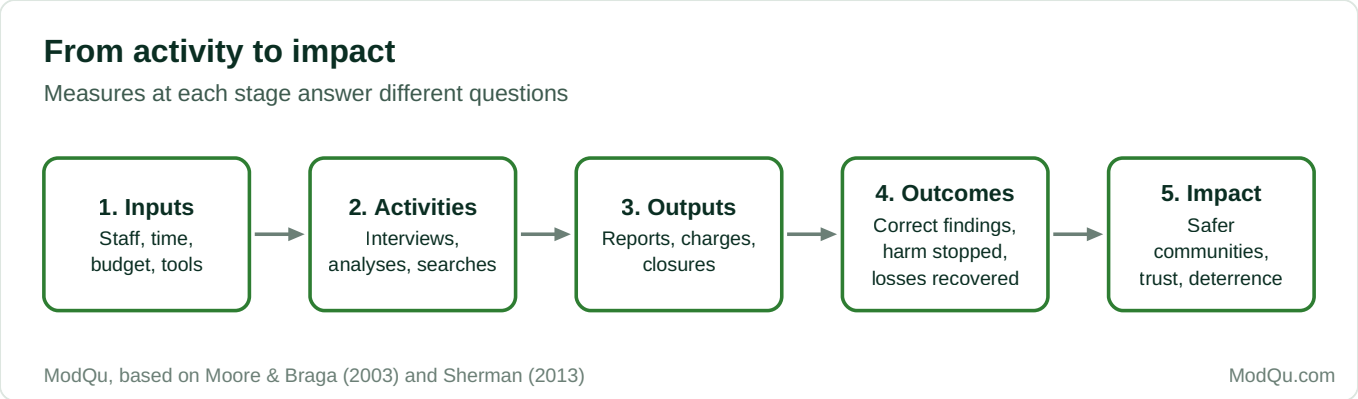


Figure 2. Clearance rates measure outputs. Effective investigation is better judged by outcomes and impact, which are harder to measure but closer to the purpose of the work. Diagram by ModQu.

Several principles help:

- **Measure outcomes, not just outputs.** Count harm stopped and correct findings, not just cases closed.
- **Adjust for case mix.** Compare like with like, taking account of the difficulty and seriousness of cases.
- **Use several measures together.** Avoid relying on any single number.
- **Audit the data.** Check that cases are recorded and classified correctly, because measures invite manipulation.
- **Include qualitative review.** Peer review of a sample of files reveals quality that numbers cannot.
- **Ask the people affected.** Survey victims, witnesses and, where possible, the people investigated.
- **Do not punish honest inconclusive outcomes.** Recording a case as not proved should not count as failure if the work was sound.
- **Review the measures.** Check periodically whether measures are driving the intended behaviour.

Measuring in non-police settings

The same issues arise for regulators, companies and other organisations that investigate. A compliance team measured only by the number of investigations closed may rush them; one measured by the number of substantiated findings may be tempted to substantiate. Better measures include the losses recovered or prevented, the time from report to decision, whether findings were upheld when challenged, whether corrective actions were implemented and worked, whether reporters felt fairly treated and whether they would report again. Occupational fraud research shows that the length of time a fraud runs before detection is strongly related to the losses it causes (Association of Certified Fraud Examiners [ACFE], 2024), suggesting that speed of detection is a meaningful outcome measure for fraud programmes.

A short example

The following hypothetical case illustrates a shift in measurement. A survey company's fraud team has been measured by the number of fraudulent respondent accounts closed each month. The number rises steadily, but client complaints about data quality do not fall. A review finds that the team has focused on easily detected individual accounts while larger, organised networks have continued to operate, and that some genuine respondents have been wrongly closed, damaging panel trust.

The company replaces the single measure with a small scorecard: the value of incentive payments to fraudulent accounts, weighted towards organised networks; the proportion of closures overturned on appeal; the time from first alert to decision; the results of quarterly peer reviews of case files; and client data quality scores. Within six months, the team closes fewer accounts but disrupts two large networks, wrongful closures fall, and client complaints decline. The new measures reward what the company actually wants.

Conclusion

Clearance rates measure whether cases are closed, not whether investigations found the truth, treated people fairly, prevented harm or earned public trust. When used as targets, they can distort practice through case selection, premature closure, misrecording and pressure for weak charges. A broader approach measures accuracy, harm reduction, fairness, timeliness, quality of process, confidence and learning, adjusts for case mix, audits its data and includes qualitative review. Investigative organisations that measure what matters, rather than what is easy to count, are more likely to achieve what investigations are for.

References

Association of Certified Fraud Examiners. (2024). *Occupational fraud 2024: A report to the nations*.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration, 84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Braga, A. A., Flynn, E. A., Kelling, G. L., & Cole, C. M. (2011). *Moving the work of criminal investigators towards crime control* (New Perspectives in Policing). Harvard Kennedy School and National Institute of Justice.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-X

Eck, J. E. (1983). *Solving crimes: The investigation of burglary and robbery*. Police Executive Research Forum.

Eterno, J. A., & Silverman, E. B. (2012). *The crime numbers game: Management by manipulation*. CRC Press.

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Harvard University Press.

Greenwood, P. W., Chaiken, J. M., & Petersilia, J. (1977). *The criminal investigation process*. D. C. Heath.

Gross, S. R., O'Brien, B., Hu, C., & Kennedy, E. H. (2014). Rate of false conviction of criminal defendants who are sentenced to death. *Proceedings of the National Academy of Sciences, 111*(20), 7230–7235. https://doi.org/10.1073/pnas.1306417111

Jackson, J., Bradford, B., Stanko, B., & Hohl, K. (2013). *Just authority? Trust in the police in England and Wales*. Routledge.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review, 70*(1), 71–79.

Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal, 18*(4), 769–783. https://doi.org/10.2307/255378

Moore, M. H., & Braga, A. A. (2003). Measuring and improving police performance: The lessons of Compstat and its progeny. *Policing: An International Journal of Police Strategies & Management, 26*(3), 439–453. https://doi.org/10.1108/13639510310489485

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.

Public Administration Select Committee. (2014). *Caught red-handed: Why we can't count on police recorded crime statistics* (Thirteenth Report of Session 2013–14, HC 760).

Sherman, L. W., Neyroud, P. W., & Neyroud, E. (2016). The Cambridge Crime Harm Index: Measuring total harm from crime based on sentencing guidelines. *Policing: A Journal of Policy and Practice, 10*(3), 171–183. https://doi.org/10.1093/police/paw003

Sherman, L. W. (2013). The rise of evidence-based policing: Targeting, testing, and tracking. *Crime and Justice, 42*(1), 377–451. https://doi.org/10.1086/670819

Sparrow, M. K. (2000). *The regulatory craft: Controlling risks, solving problems, and managing compliance*. Brookings Institution Press.

Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review, 37*(3), 513–548. https://doi.org/10.1111/1540-5893.3703002

Wellford, C., & Cronin, J. (1999). *An analysis of variables affecting the clearance of homicides: A multistate study*. Justice Research and Statistics Association.