vehicles

MDPI

Article

Diagnosing and Raising Road Traffic Demerit Point Enforcement Reliability

Kelvin Agabu Lungu [1], Mbulelo Phesa [2,*], Elizabeth N. Kangwa [1] and Augustine Mutale [1]

[1] Department of Civil and Environmental Engineering, School of Engineering, University of Zambia, Lusaka P.O. Box 32379, Zambia; kelvin.lungua@unza.ac.zm (K.A.L.); elizabethkangwa3113@gmail.com (E.N.K.); mutalesa95@gmail.com (A.M.)
[2] Department of Civil Engineering, iWS-CIDB Centre of Excellence, Faculty of Engineering, Built Environment and Information Technology, Walter Sisulu University, iYunivesith Walter Sisulu, Buffalo City Campus, Ku-Gompo City 5247, South Africa
* Correspondence: mphesa@wsu.ac.za; Tel.: +27-73-502-9886

Abstract

A demerit point deters only if every administrative stage from detection to sanction works. This exploratory, planning-level study diagnoses the perceived reliability of that chain and models a workflow intended to strengthen it. It applies series reliability to a survey of 384 licenced drivers in Lusaka, Zambia. Record integration was the lowest-rated stage in every bootstrap draw and under three index definitions; the chain index, a comparison scale and not a probability, was 0.09 (95% interval 0.08 to 0.10). Record integration and sanction processing were the stronger correlates of perceived deterrence in the administered scales. An exploratory sensitivity check with overlapping items removed or reallocated, which could not use respondent-level answers, kept the pair ahead in most but not all plausible item correlation structures, and the lowest stage depended on the allocation of items. The design models four control points and two backup routes with common-cause failure, using assumed effect ranges. Under these assumptions the full workflow gave a median chain index of 0.40 (90% simulation interval 0.33 to 0.49), and between 0.19 and 0.41 in thirteen alternative cases. No single stage could reach 0.25 alone, backup routes lost value as shared failures rose, and the catch rates of the unique identifier and of automatic posting carried most of the value of further information. The results describe a constructed index under stated assumptions. They are candidate priorities for administrative audit and pilot evaluation, not demonstrated improvements in enforcement, deterrence or road safety.

Keywords: certainty of sanction; common-cause failure; control points; enforcement reliability; road safety; series reliability

Check for updates

Received: 17 October 2026
Revised: 13 November 2026
Accepted: 18 November 2026
Published: 27 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Road traffic injury is a leading cause of death worldwide, and the burden falls most heavily on low- and middle-income countries, where the African region recorded a rise in road deaths while the global total fell [1]. A demerit point deters only if the administration behind it works. An offence must be detected, matched to the correct driver, recorded as points and followed by a sanction when a threshold is crossed, and failure at any step removes the deterrent value of the others. Enforcement is among the few measures that a road agency can change within a budget cycle, yet it is weakly applied: a review of 48 African countries rated the effectiveness of speed and drink-driving enforcement at 3.9 and 3.5 out of 10, and found fragmented records and lead agencies that often lacked

the mandate, capacity or funding to lead [2]. Reviews report that traffic laws are rarely enforced in several African countries and that legislation works best when paired with strong enforcement [3,4]. Points-based licencing is spreading nonetheless: South Africa legislated administrative adjudication of road traffic offences in 1998 and amended it in 2019 to improve the service of documents on infringers [5], Zambia gave its demerit point system a legal basis in 2024 [6], and Kenya announced in 2026 a digital framework for minor offences under which notices can be delivered by text message or email, including for offences detected by traffic cameras [7].

Evaluations of points systems ask whether outcomes change. A worldwide meta-analysis found that the systems are usually associated with fewer violations and casualties, with effects that depend on the context and on how well they are run [8]; in Spain the effect faded within a few years [9], and a regression discontinuity analysis found deterrent effects in Italy [10]. Deterrence theory explains why: the perceived certainty and swiftness of a sanction matter more than its severity [11,12], and perceived fairness determines whether a sanction is accepted [13]. Certainty is therefore a property of the whole administrative chain that police, licencing and adjudication bodies operate, and not of the statute alone.

Reliability engineering has a vocabulary for such a chain. A process that runs through stages in sequence is a series system, and its reliability is the product of the stage reliabilities [14], which is lower than that of the weakest stage. Reliability can be raised by control points, which compare the output of a stage with an independent reference and correct a discrepancy, or, by redundancy, a second path that can succeed when the first fails. Layered defences explain why several imperfect layers can protect where one cannot [15], but common-cause failure limits redundancy because routes that depend on the same record, system or staff fail together [14]. The data on which a chain depends are often fragmented: only 18 of 48 African countries used the standard 30-day definition of a road death [2], and a capture–recapture study in Zambia found low completeness in each of three sources used alone [16]. Survey ratings are ordinal perceptions, so an index built from them has no natural probability scale, and the stages are not independent. The effects of administrative controls on such a chain are also unmeasured in most low- and middle-income settings.

To our knowledge no study diagnoses which stage of a demerit point chain limits it, with the uncertainty of that diagnosis, and then models what a workflow of controls could add. Studies of points systems estimate whether outcomes change [8–10] and do not decompose the chain behind them. An agency with a fixed budget needs both halves: a diagnosis shows where to act, and a design shows what to do there and which assumptions to measure before spending, because spending on the most visible stage, such as officers or cameras, adds little if offences cannot be linked to the right driver or sanctions arrive late.

The problem is not specific to one country. The worldwide review found that the results of points systems depend on how well they are administered [8]. As detection is increasingly automated, and reviews of speed cameras report fewer crashes where they are used [17], the point of failure moves to the stages that follow detection. Fragmented records, enforcement shared among several agencies and uneven capacity to follow an offence through to a sanction are described across low- and middle-income countries [2,18–20], and in Africa the effect of enforcement depends on the administrative and legal setting [3]. The same stages appear in the points systems of South Africa, Kenya and the European Union, and the diagnosis, the model and the simulation apply to any enforcement chain whose stages can be rated or measured. What is local is the set of stage values. Lusaka, the capital of Zambia, was chosen as an early-stage case [21,22]: it holds the national road transport regulator and in 2022 accounted for 18,393 of the 33,682 crashes recorded nationally (54.6%), and the system is new, so the ratings describe its first administration, when correction costs

least. The case is a single city, and its stage values are not claimed to represent other places; what it supports is the transfer of the method.

This paper has two objectives that share one baseline. The first is to diagnose the perceived reliability of the enforcement chain behind the demerit point system in Lusaka: to describe how drivers rate each stage and the deterrent outcome, to estimate the association and relative importance of each stage with perceived deterrence, to compute a chain index with its uncertainty, to test whether the lowest stage is stable under different index definitions and dependence, and to compare sequences of repair. The second is to design a workflow of four control points and two backup routes, each with an owner and an indicator, and to estimate under stated assumptions the gain of each element and combination, the limits set by common-cause failure and by a single stage, and the value of further information. The paper contributes to the literature on certainty of sanction and the administration of criminal justice a way to apply series reliability to ordinal perception data, a ranking of stages reported with the uncertainty of the ranks, a model of what controls could add that states its assumptions and shows which of them to measure, and a measurement plan and implementation framework for police, licencing and adjudication institutions.

2. Materials and Methods

2.1. Design and Policy Context

The study has two parts that share one baseline. The diagnosis is a cross-sectional survey of drivers' perceptions of each stage of the enforcement chain, supported by key informant interviews. The design is a model-based scenario study that starts from the stage indices of the diagnosis. All indices are planning scales built from perceived ratings and not measured probabilities, the effects of the controls are assumed, and the analysis is exploratory and describes associations and conditional results, not causes. Throughout, a perception-derived quantity is called an index, and one minus an index is called its shortfall; neither is a measured probability of enforcement.

Road safety policy sets the frame within which the chain operates. The United Nations General Assembly proclaimed a Decade of Action for Road Safety for 2021 to 2030 with a target of halving road deaths and injuries [23], and its global plan calls for strong enforcement managed by a lead agency [24]. The African Road Safety Charter commits member states of the African Union to road safety management, enforcement and data systems [25]. In Zambia the Road Traffic Act No. 11 of 2002, as amended by the Road Traffic (Amendment) Act No. 8 of 2022, defines offences and licencing [26], the Road Transport and Safety Agency Act No. 13 of 2002 establishes the regulator that administers them [27], and the Road Traffic (Demerit Point System) Regulations, Statutory Instrument (SI) No. 53 of 2024, added the demerit point system [6]. The regulator announced countrywide sensitisation before implementation [28] and held a sensitisation exercise for Members of Parliament in July 2025 [29], so the ratings in this study describe the chain during its roll-out and not a settled routine. The Data Protection Act No. 3 of 2021 and the Electronic Communications and Transactions Act No. 4 of 2021 govern personal data and electronic records [30,31], so the linking of offences to driver records is where road safety policy and data protection law meet. The International Organization for Standardization (ISO) standard 39001 requires an organisation to set road traffic safety performance factors, to monitor them and to review them [32].

The legal process places further steps inside the chain. Points attach to a driver through the adjudication route that the Act and the Regulations prescribe, whether a notice is settled or an offence is adjudicated and a conviction recorded; a driver may contest a notice or a decision; and points accumulate across offences until the maximum is passed and a licence action follows [6,26,28]. The three-stage structure assigns these steps to the stage where

they occur. Detection is the sensing and capture of the offence. Record integration is the attribution of the offence to the correct vehicle owner and driver and the accuracy and consistency of the record. Sanction processing is every step from the posting of points to the consequence, namely the notice, settlement or adjudication, contest, accumulation to the threshold and licence action. Contest and appeal are represented by the fairness items of the acceptance scale and by the driver check of the balance (E6), and accumulation is represented by the fact that a shortfall at any offence leaves the balance incomplete, so that a threshold that needs several recorded offences is reached less often than a single offence is sanctioned. Accumulation therefore increases, and does not remove, the weight of the weakest stage.

2.2. Setting, Sample and Instruments

The population was licenced drivers who held a valid Zambian driving licence and had driven in Lusaka for at least twelve months. Fieldwork took place from July to September 2026. Because no deduplicated count of licenced drivers exists, the population was treated as large, and Cochran's formula for a proportion [33] at 95% confidence was used; a proportion of 0.5 and a margin of 0.05 gave 384 completed questionnaires; 420 were fielded to allow for non-response. The sample was stratified by vehicle and licence class, as Table 1 shows. The stratum weights were set before fieldwork from the composition of the Lusaka vehicle fleet and licencing activity reported by the regulator, with the public service and commercial classes weighted up to reflect their greater exposure to enforcement. No register of licenced drivers by class exists, so the weights are planning allocations and not population shares, and the results are reported without reweighting. Within each stratum, respondents were approached at high-throughput sites that the class concerned uses, such as vehicle inspection stations, licencing offices, transport termini, fuel stations and major junctions, and every nth eligible driver at a site was invited to take part. Questionnaires were self-administered and anonymous. Of 420 questionnaires, 391 were returned (93.1%) and 384 were complete and usable (91.4%); the seven returned questionnaires that were incomplete or invalid were excluded, no missing values were imputed, and the usable questionnaires followed the allocation in Table 1. Section A of the questionnaire recorded vehicle class, gender, age group, years of driving in Lusaka, driving frequency, weekly distance and whether the respondent had received demerit points or a licence-related action; these characteristics are reported in Section 3.1. This was not a probability sample, so the intervals reported for the survey describe sampling variability under random selection within sites, are indicative, and are not population confidence intervals.

Table 1. Stratified allocation of the driver sample.

Stratum (Vehicle and Licence Class)	Stratum Weight	Required Sample	Fielded Target
Private light-vehicle drivers	0.55	211	231
Public service vehicle drivers (taxi and minibus)	0.22	85	92
Commercial and heavy goods vehicle drivers	0.13	50	55
Motorcycle riders	0.10	38	42
Total	1.00	384	420

Five scales of five items were scored from 1 (strongly disagree) to 5 (strongly agree), so that a higher score meant a more reliable stage or a stronger deterrent effect; no item was reverse scored and no "do not know" option was offered. The scales measured detection (certainty and coverage), record integration (accuracy, timeliness and consistency of records), sanction processing (speed, uniformity and follow-through), acceptance (perceived respect,

neutrality, explanation and fairness) and deterrence (change in behaviour and credibility of the chain). Table 7 gives the wording of all 25 items and the scale to which each was assigned. Detection and acceptance items were adapted from enforcement-perception and procedural-justice instruments, record integration and sanction processing items were written from the operational definition of each stage, and all items were checked by expert review and piloted. A stage score is the mean of its five items, and internal consistency was assessed with Cronbach's alpha [34]. The scales measure perception. Respondents observe some stages, such as the stops they meet on the road, more directly than others, such as record reconciliation, and no comparison with administrative records was available, so the scales were not validated against measured performance. Two features of the wording matter for interpretation. The detection item on the predictability of checks can be read as greater coverage or as an opening for evasion. Two record integration items, on the timely posting of points and on the checking of the point balance, describe functions that the design places in sanction processing (Section 2.8). Sections 2.5 and 3.2 therefore report checks that remove or reallocate these items.

2.3. Key Informant Interviews

Six key informants working in detection, licencing and records and in road safety management were recruited by snowball sampling [35], which began with an engagement with the Central Police Department and continued by referral until each stage of the chain was represented. They were interviewed with verbal consent, using a semi-structured guide that took each informant through the operation of the stage in which they worked and the points at which offences fail to be detected, recorded, linked or sanctioned. Interviews were audio-recorded, coded by stage and by failure mode, and reported as paraphrases attributed to informant numbers (KI 1 to KI 6) without roles. The accounts are explanatory material and not a validation of the indices or of the model: each failure mode that an informant described was linked to the survey statement that addresses it, so that the ratings could show whether the account was supported, and the links that the ratings did not support are reported with those that they did.

2.4. Stage Indices and the Chain Index

Figure 1 shows the structure of the diagnosis. Detection senses the offence, record integration attributes it to a driver and updates the points on the record, and sanction processing converts the accumulated points into a consequence. Stage scores are the means of their five items. Treating such means as continuous is supported by simulation evidence [36,37]. For a stage with mean score m, the index of Equation (1) maps the lowest score to 0, the neutral score of 3 to 0.5 and the highest to 1,

$$r = \frac{m - 1}{4} \tag{1}$$

and the chain index for the three technical stages, detection, record integration and sanction processing, is the product of the stage indices (Equation (2)),

$$R = r_{\text{det}}\, r_{\text{int}}\, r_{\text{san}} \tag{2}$$

where r is the stage index in the diagnosis and the stage reliability in the design. The index is a common scale for comparing stages and designs and not an estimate of the share of offences that end in a sanction. Acceptance, the perceived fairness and legitimacy of the process, and perceived deterrence are not technical stages and enter only the regression, and a variant that adds acceptance as a fourth factor was computed as a check. In this paper an index is a rescaled mean rating, one minus an index is its shortfall, and the word reliability in the design refers to an index on the same scale. Statements in the language of

probability are conditional on reading the index as a probability, which the survey does not support.

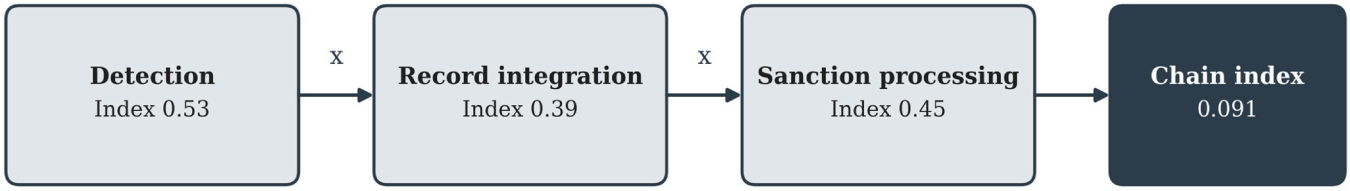

Figure 1. Structure of the enforcement chain with baseline stage indices.

2.5. Sampling Uncertainty and Sensitivity Checks

The sampling uncertainty of the indices came from a parametric bootstrap of 200,000 draws of the five stage means from a multivariate normal distribution with the observed means, standard errors and correlations, which also gives the probability that each technical stage is the lowest. Discriminant validity was checked by correlating the scales after correction for unreliability, with values well below 0.85 taken as evidence that the scales are distinct [38]. The bootstrap and t intervals describe the sampling variability of the sample estimates under random selection within sites and strata; because recruitment was not a probability sample they are indicative and are not population confidence intervals. Correlation matrices were resampled by drawing 384 observations from a multivariate normal distribution with the covariance of the scale scores and recomputing the correlations. Calculations used Python 3 with NumPy and fixed seeds: 11 for the 200,000 draws of the stage means and 3 for the resamples of the alternative allocations of Tables 5 and 8 for the 4000 resamples of the dominance analysis, 20,260 for the 20,000 resamples of the path model, and 2026 and 2027 for the families of item correlation matrices described below. The code is provided as Supplementary Materials. The intervals of the design are simulation intervals that propagate assumed ranges and are conditional on those ranges.

Two further checks concern what the product means. First, the index was also defined as the share of respondents who agree with the items of a scale and as the share who do not disagree; if the lowest stage is the same under all three definitions, the diagnosis does not depend on the definition. Second, if the index is read as a probability and the stages are positively associated, the probability that all stages succeed is at least the product [39] and cannot exceed the lowest stage probability (Equation (3)),

$$r_{det}\, r_{int}\, r_{san} \;\leq\; P(\textit{all stages succeed}) \;\leq\; min(r_{det}, r_{int}, r_{san}) \tag{3}$$

An illustrative value was obtained with a Gaussian copula [40] that has the stage indices as margins and the observed correlations between the three scales as dependence. Those correlations are between respondents' ratings of different stages and show how strongly the ratings move together; they do not measure whether the same offence succeeds or fails at two stages, so the copula values are dependence scenarios for illustration and not estimates of the chain index.

Three exploratory sensitivity checks concern the items. The first removed the items that overlap in content across stages: the detection item on recording the offence against the driver, the record integration items on the timely posting of points and on checking the point balance, and the sanction processing item on the speed of assigning points, and from deterrence the items on repeated offending and on the credibility of the chain, which overlap with the stage items. The second reallocated items to the stage that the design assigns them to: the two record integration items on posting and on the balance went to sanction processing and the detection item on recording the offence against the driver went to record integration. The third removed the detection item on the predictability of checks.

Each check was applied to the stage indices (Table 8) and, for the first two, to the regression and dominance analysis. The respondent-level answers cannot be located, so the variances and covariances of the altered scales could not be computed from them (their means follow exactly from the item counts), and the main results do not use these checks. The covariances of the altered scales were derived instead from the item response counts (Table 7), the scale correlations (Table 6) and Cronbach's alpha. These fix the item means and standard deviations, the reliability of each scale and the correlations between the unaltered scales, but not the correlations between individual items. To show what this leaves open, the regression, dominance analysis and resampling were repeated over 320 item correlation matrices. Each reproduces the reported alphas and scale correlations to within 0.004 and differs in how strongly individual items correlate. The item correlations were perturbed around the common-correlation matrix with standard deviations of 0.04, 0.08 or 0.10 (0.04, 0.06 or 0.07 between scales), the matrix was adjusted to the nearest valid correlation matrix, and it was kept only if it met the tolerance. Each matrix was resampled 300 times. Results are reported as the median and the range across matrices and are read as a sensitivity analysis, not as estimates.

2.6. Regression, Dominance and Path Analysis

Perceived deterrence was regressed on the four stage scores (Equation (4)),

$$Y = b_0 + b_1 X_{det} + b_2 X_{int} + b_3 X_{san} + b_4 X_{acc} + \varepsilon \quad (4)$$

where Y is the deterrence score and the X terms are the stage scores. Standardised coefficients, variance inflation factors and the shrinkage-corrected coefficient of determination (Stein formula for cross-validity) [41] were computed from the correlation matrix and the sample size. Relative importance was assessed by dominance analysis [42,43], which attributes to each predictor j its average increment to the coefficient of determination over all subsets S of the other predictors (Equation (5)),

$$C_j = \sum_{k=0}^{3} \frac{1}{4\binom{3}{k}} \sum_{S \subseteq \{1,\ldots,4\} \setminus \{j\}, |S|=k} \left[R^2(S \cup \{j\}) - R^2(S) \right] \quad (5)$$

with Johnson's relative weights as a second estimator [43,44]. The uncertainty of the ranks came from 4000 resamples of the correlation matrix, each drawn as 384 observations from a multivariate normal distribution with the observed correlations. Because the chain is a sequence, a recursive path model was also fitted from the correlation matrix: record integration was regressed on detection, sanction processing on detection and record integration, and deterrence on the four stage scores. The indirect effect of a stage is the sum, over the paths through later stages, of the products of the coefficients along each path, the total effect is the direct and indirect effects together, and 95% intervals came from 20,000 resamples of the correlation matrix drawn in the same way. The order of the stages is taken from the process and is not tested. The regression and the total effects were also used to express the difference in the deterrence score that is associated with a stage differing by half a scale point, by the gap to a score of 3.5 (halfway between neutral and agree) or by the gap to the mean rating of detection, with the gap between mean ratings treated as fixed. These associated differences are statistics of one cross-sectional questionnaire and are not predicted effects of an intervention. Because the stage ratings and the deterrence ratings came from the same respondents in one questionnaire, common-method variance and conceptual overlap between items may inflate the associations [45], and the regression was repeated with overlapping items removed (Section 2.5).

2.7. Repair Scenarios and Spending Rules

For repair, the sensitivity of the chain index to a stage is the product of the other two stage indices (Equation (6)),

$$\frac{\partial R}{\partial r_i} = \prod_{j \neq i} r_j, \qquad \frac{\Delta R}{R} = \frac{\Delta r_i}{r_i}, \tag{6}$$

so the proportional gain from a given rise is inversely proportional to the current stage index. Scenarios raised stages by 0.10 or to a stated level. Three rules for spending a budget of index points at equal cost were compared: raise the weakest stage first, spread the budget evenly, and concentrate it on the strongest stage. The first rule maximises the product for a given total, because the product of numbers with a fixed sum is greatest when they are equal. The comparison assumes that one index point costs the same in every stage.

2.8. Workflow Design

Each stage was given at least one control point, and the two stages that depend on identifying and reaching a person, record integration and sanction processing, also received a backup route. Each element has an owner function and an indicator computed from system records, as ISO 39001 expects [32]. Figure 2 shows the six elements, with control points above the stages and backup routes below them, and Table 2 gives the mechanism, owner and indicator of each. E1 monitors the availability of detection equipment. E2 is a unique identifier captured at the offence and checked against the licence register, and E3 is a second key, such as the chassis number or national identity number, to match the vehicle to the owner. E4 posts points automatically with daily reconciliation against the offence register, E6 lets drivers check their points balance and challenge errors, and E5 is a second notice channel, such as a text message besides the postal notice. No element requires a new statute, although those that use additional personal identifiers must meet tests of necessity and proportionality [30].

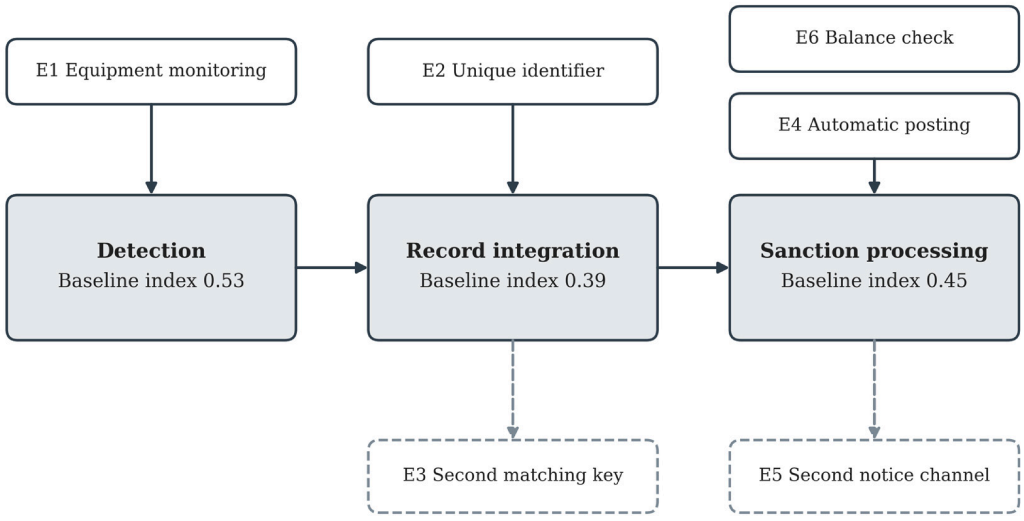

Figure 2. Workflow of control points and backup routes, with baseline stage indices.

The elements act on different failure modes, and an element recovers a failure only if it acts on that mode. E1 addresses offences that pass unrecorded because a device is out of service. E2 addresses offences that are attributed to no driver or to the wrong driver: it captures an identifier of the driver, such as the licence number where the driver is stopped or on a verified declaration where the offence is recorded against a vehicle, and checks it against the licence register. E3 addresses vehicle records for which the plate does not match

an owner, by a second key. Identifying the owner of a vehicle is not treated as identifying the offending driver: a notice addressed to an owner leads to points for a driver only through the nomination or adjudication route that the law prescribes. Database matching is the comparison of the offence record with the licence and vehicle registers by E2 and E3 and the daily reconciliation of E4. E4 addresses points that are not posted, are posted late or are posted twice. E5 addresses notices that do not reach the person, and a notice is counted as served only when delivery is recorded. E6 addresses wrong or missing points: a challenged point is flagged, does not count toward threshold action until the licencing unit has resolved the challenge within the set period, and is corrected through the record. An incorrect match or a disputed point is therefore resolved by E3, E4 and E6 with a recorded outcome, and the unresolved share is carried in the indicator of E6.

Table 2. Workflow elements with owner and indicator.

Element	Stage and Type	Mechanism and Trigger	Owner	Indicator
E1 Equipment availability monitoring	Detection; control point	Automatic alert when a device has sent no offence record for a set period; repair dispatched	Regulator operations unit	Share of device-hours in service
E2 Unique identifier at offence and licencing	Record integration; control point	Identifier of the driver or owner captured at the offence and checked against the licence register before the record is accepted	Licencing unit	Share of offences linked to a verified driver record
E3 Second matching key for the owner record	Record integration; backup route	If the plate does not match an owner, the chassis number or the owner identity number is tried	Vehicle registry	Share of unmatched offences resolved by the second key
E4 Automatic posting with daily reconciliation	Sanction processing; control point	Points posted by the system without manual entry, and a daily comparison of offences and points flags differences	Licencing unit	Share of offences with points posted within one day
E5 Second notice channel	Sanction processing; backup route	If the postal notice is not confirmed as delivered, a text message or portal notice is sent	Regulator enforcement unit	Share of notices confirmed as received
E6 Driver check of the points balance	Sanction processing; control point	Drivers query their balance and challenge an error, which the licencing unit corrects	Licencing unit	Share of challenges resolved within a set period

2.9. Reliability Model and Propositions

The baseline of the design is the diagnosis: the stage indices in Table 3, which come from the survey, are used as stage reliabilities and their product as the baseline chain index. The standard error of each stage index was about 0.01, and baseline uncertainty was represented in the simulation by a normal distribution with that standard deviation around each index. The model is a model of an index. If the index were read as a probability it would be a conventional reliability model, but it is used here to describe how a constructed index would change under assumed recoveries, and its results are not estimates of the share of offences that end in a sanction.

Table 3. Baseline ratings and indices of the stages.

Stage	Mean Rating (1 to 5)	Index	Shortfall (1 Minus Index)
Detection	3.12	0.530	0.470
Record integration	2.54	0.385	0.615
Sanction processing	2.78	0.445	0.555
Chain (product)		0.091	0.909

Let r be the reliability of a stage, an index on the scale of Equation (1), and f = 1 − r its shortfall. A control point detects a failure with probability d, the catch rate, and corrects it with probability g, the correction rate, so that the reliability becomes (Equation (7))

$$r' = r + (1 - r)\, d\, g \tag{7}$$

Control points in one stage act in turn on the failures that remain. A backup route has its own reliability r_{b} and is not independent of the main route. The beta-factor model of common-cause failure [14] assigns a share β of the failures of the main route to causes that the backup route shares, such as the same record, system or staff, so that (Equation (8))

$$f' = f\,[\beta + (1 - \beta)(1 - r_{\mathrm{b}})], \qquad r' = 1 - f' \tag{8}$$

When β is zero the routes fail independently, and when it is one the backup route adds nothing. Control points were applied before backup routes within a stage, and stage reliabilities multiply as in Equation (2). Four propositions follow, each a property of the model and not evidence about real systems. Proposition 1—the gain of a stage depends on the other stages. Recovering a share s of the failures of stage i adds (Equation (9))

$$\Delta R_i = s\,(1 - r_i)\prod_{j \neq i} r_j = s\,R\,\frac{1 - r_i}{r_i} \tag{9}$$

so for the same share recovered the gain is largest at the weakest stage. Proposition 2—redundancy has a ceiling set by common cause. The reliability of a stage with a backup route cannot exceed $1 - \beta f$, and the gain from the route, $f(1 - \beta)r_{\mathrm{b}}$, falls linearly in β. Proposition 3—a single stage has a ceiling. With the other stages unchanged, the chain index cannot exceed the product of their reliabilities, so a target T can be reached by repairing stage i alone only if T does not exceed that product, and the share of its failures to recover is (T divided by the product of the other reliabilities, minus r_{i}) divided by $(1 - r_{\mathrm{i}})$. Proposition 4—priority depends on the weakest stage and on effort. A later element on the same stage adds less than it would alone, because fewer failures are left for it, so the ranking changes as stages are repaired.

2.10. Assumed Effects and Evidence

The effects of the six elements have not been measured on the administrative stages of a points system, so the ranges in Table 4 are assumptions set from the failure mechanisms of each stage, made wide and drawn uniformly. Effort tiers, from 1 for a procedural change to 3 for a new integration or a change in legal basis, are also assumptions and only order the elements. Published evidence, indirect in every case and drawn from other domains, was used to check that the ranges are not implausible and not to calibrate them. For example, the recovery assumed for E2 and E4, a catch rate times a correction rate of 0.15 to 0.63, lies below the 81% reduction reported for computerised order entry. No comparable measurement was found for E1 or for the common-cause share of any backup route. The catch rate and the correction rate are defined over all the failures of the stage, so their product is the share of the failures of the stage that the element recovers; an element that acts on only some

failure modes has a smaller effective recovery, which Section 2.11 tests. Other distributions, correlated parameters and less optimistic ranges are also tested there.

Table 4. Assumed ranges and indirect published evidence.

Element	Assumed Ranges	Effort Tier	Published Evidence
E1 Equipment monitoring	Catch 0.30 to 0.70; Correction 0.50 to 0.90	2	None found
E2 Unique identifier	Catch 0.30 to 0.70; Correction 0.50 to 0.90	3	Linking police and hospital crash records was projected to cut age errors from 15.7% to 0.2% [46]
E3 Second matching key	Backup reliability 0.30 to 0.60; Common-cause share 0.20 to 0.60	3	Adding date of birth to name cut the share of hospital records sharing a match key from 16.49% to 40.66% to 0.16% to 15.47% across five institutions [47]
E4 Automatic posting	Catch 0.30 to 0.70; Correction 0.50 to 0.90	3	Computerised order entry cut non-missed-dose medication errors by 81% [48]
E5 Second notice channel	Backup reliability 0.40 to 0.70; Common-cause share 0.20 to 0.50	2	Text reminders raised average payment by 145% in England [49]; text and email outperformed letters in Rwanda [50]; none for common cause
E6 Balance check	Catch 0.20 to 0.50; Correction 0.50 to 0.90	1	21% of 22,889 patients who read their visit notes reported a mistake [51]

2.11. Simulation and Value of Information

The baseline indices of the diagnosis were drawn with a standard deviation of 0.01 to reflect sampling error, and 200,000 draws of all parameters at once supported seven analyses: four packages (control points only, backup routes only, both, and the two record integration elements alone); the gain of each element on its own and the probability that a package reaches a target chain index; the order of introduction by gain per unit of effort and by gain alone; the effect of varying β from 0 to 1 with other parameters at the middle of their ranges and the share of failures of a single stage that must be recovered to reach a target; the sensitivity of the full package by Spearman rank correlation; a comparison of the gains predicted by Equation (9) with the simulated medians; and a value-of-information analysis of the choice of the first element to introduce. For the last, the expected value of perfect information was the mean over draws of the best element in each draw minus the best element on average, and the share carried by each parameter was estimated by dividing the draws into twenty equal groups by that parameter [52], for a choice by gain alone and by gain per unit of effort. Each result was checked by an independent recomputation.

Sixteen cases, the assumed case and fifteen alternatives, tested whether the priorities depend on the assumptions (Table 25): the assumed ranges; triangular distributions with the mode at the middle or at the lower quarter of each range; parameters correlated through a Gaussian copula with a common correlation of 0.5 or 0.8; the less optimistic half of every range (the lower half for catch rates, correction rates and backup reliability and the upper half for common-cause shares); recovery and backup reliability halved; each control addressing 30% to 70% of the failures of its stage; a common-cause share of 0.5 to 0.9; the last three less optimistic conditions together; no benefit from E2 and E4; and baselines taken from the alternative item allocations and index definitions of Table 8. The order of introduction was recomputed with effort weights of 1, 2 and 3 for the three tiers as assumed, with weights of 1, 2 and 4, 1, 2 and 6 and 1, 3 and 9, and with equal weights, which gives the order by gain alone (Table 26). Each analysis used 200,000 draws with fixed seeds. In each draw the baseline indices and every parameter were sampled, the reliability of each

stage was computed from Equation (7) for the control points in turn and then Equation (8) for the backup route, and the chain index from Equation (2).

2.12. Use of Generative AI

During the preparation of this study, Claude (Anthropic, Version 2.9939.2, build d3e504) was used to assist in developing scripts for statistical analyses, simulations, and for graphical representations of Figure 15 (implementation framework) using the results from this manuscript. Furthermore, Claude was also used to verify the internal consistency of reported results and refine the manuscript's language and clarity. All AI-assisted outputs were critically reviewed and validated by the authors, who take full responsibility for the accuracy, integrity, reliability and scientific validity of all AI-generated content and the final manuscript.

3. Results

3.1. Diagnosis: Ratings of the Stages

The 384 usable questionnaires matched the planned total (Table 1). All five scales were internally consistent, with Cronbach's alpha from 0.855 to 0.883 (Table 5). Table 5 reports scores from 1 to 5, with the disagree and agree shares averaged over the five items of each scale and 95% intervals from the t distribution. Detection, with a mean of 3.12 (95% interval 3.04 to 3.20), was the only stage whose interval lay wholly above the neutral score of 3, and acceptance (3.02, 2.94 to 3.10) did not differ clearly from it (Figure 3). Deterrence (2.88), sanction processing (2.78) and record integration (2.54) were below it. Only 15% of drivers agreed with the record integration items on average, against 36% for detection, and 48% disagreed. Record integration was 0.24 points below sanction processing (standard error 0.04, from the paired ratings), and sanction processing 0.34 points below detection (standard error 0.04), both paired differences exceeding five standard errors.

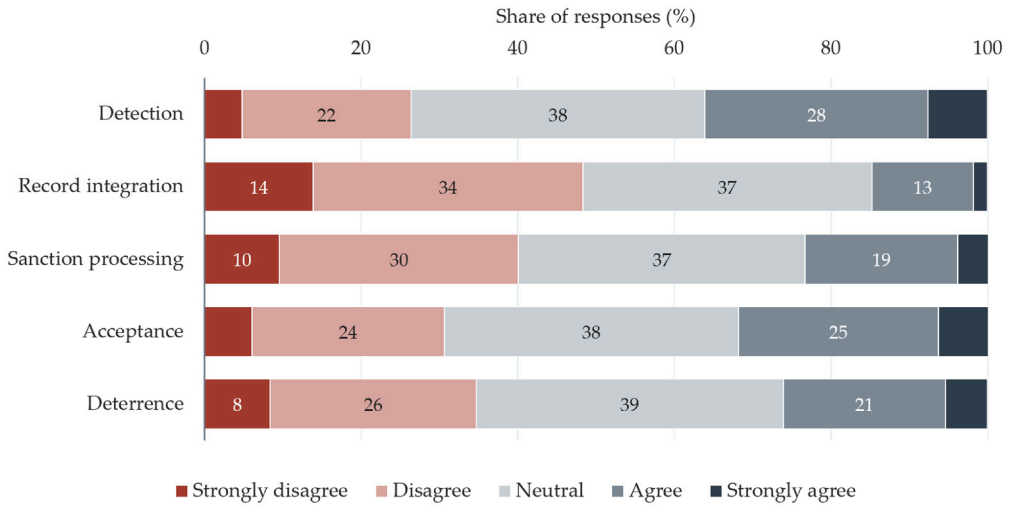

Figure 3. Distribution of responses to the five scales.

Respondents were mostly men (332, 86.5%; 52 women) and were aged 26 to 35 years (136) or 36 to 45 years (103); 59 were aged 18 to 25, 62 were aged 46 to 55 and 24 were aged 56 or above. Years of driving in Lusaka were 1 to 3 for 71 respondents, 4 to 7 for 131, 8 to 15 for 120 and more than 15 for 62. Most drove daily (251) or several times a week (76), and 78, 153, 103 and 50 respondents drove under 100 km, 100 to 300 km, 301 to 600 km and more than 600 km a week. Prior receipt of demerit points or a licence-related action was reported by 168 respondents (43.8%), denied by 174 (45.3%) and uncertain for 42 (10.9%). Exposure

to enforcement on the road was therefore high and direct experience of the points system was common, but ratings of stages that a driver cannot observe, such as the reconciliation of records, rest on indirect knowledge, and the sample was predominantly male, as drivers in the public service and commercial classes are.

Table 5. Scale statistics, $n = 384$.

Scale	Mean	SD	95% Interval of the Mean	Alpha	Disagree (%)	Agree (%)
Detection	3.12	0.80	3.04 to 3.20	0.873	26	36
Record integration	2.54	0.74	2.47 to 2.61	0.855	48	15
Sanction processing	2.78	0.80	2.70 to 2.86	0.868	40	23
Acceptance	3.02	0.82	2.94 to 3.10	0.883	31	32
Deterrence	2.88	0.80	2.80 to 2.96	0.869	35	26

Figure 3 shows the response shares, with segments below 8% unlabelled; record integration has the largest share of disagreement and detection the largest share of agreement. Table 6 gives the correlations between the scales. Observed correlations ranged from 0.30 to 0.58, and after correction for unreliability from 0.34 to 0.66, well below the 0.85 at which scales cannot be told apart [38], so the scales were treated as distinct. The strongest observed correlation with deterrence was that of sanction processing (0.58) and the weakest that of acceptance (0.45). In the table, the observed correlations are below the diagonal, the correlations corrected for unreliability above it, and Cronbach's alpha on the diagonal.

Table 6. Correlations between the five scales.

	Detection	Integration	Sanction	Acceptance	Deterrence
Detection	(0.873)	0.453	0.494	0.395	0.564
Integration	0.391	(0.855)	0.521	0.342	0.624
Sanction	0.430	0.449	(0.868)	0.455	0.662
Acceptance	0.347	0.297	0.398	(0.883)	0.518
Deterrence	0.491	0.538	0.575	0.454	(0.869)

Table 7 gives the wording and the response counts of each of the 25 items, scored from 1 (strongly disagree) to 5 (strongly agree). The scale means recomputed from the item counts agreed with the scale scores to within 0.01 for all five scales. At the item level the lowest mean was for the statement that a driver can easily verify a current point balance (2.36), with which 10% agreed and 56% disagreed, followed by the timely updating of points (2.42). The highest was the likelihood that offences are detected (3.30, 42% agreeing). All five record integration items had means of 2.72 or lower, and more respondents disagreed than agreed on each. Four of the six lowest-rated statements belonged to record integration and two to sanction processing, and none to detection or acceptance. The lowest-rated items of sanction processing were uniform treatment of comparable cases (2.54) and the speed of licence action (2.69), and the lowest-rated item of acceptance was the explanation of reasons (2.78).

Table 7. Items and response counts, $n = 384$.

Scale	Item	1	2	3	4	5	Mean	SD
Detection	Traffic offences committed in Lusaka are very likely to be detected by enforcement.	8	68	147	122	39	3.30	0.95

Table 7. *Cont.*

Scale	Item	1	2	3	4	5	Mean	SD
	Enforcement coverage (patrols, cameras, checks) across the roads I use is high.	25	95	144	95	25	3.00	1.01
	The timing and location of enforcement checks are predictable enough to be relied upon.	28	109	134	96	17	2.91	1.00
	It is difficult for a driver to evade detection when committing an offence.	13	79	132	130	30	3.22	0.97
	When an offence is detected, it is reliably recorded against the driver responsible.	18	64	163	104	35	3.19	0.98
Record integration	Offence records are accurately linked to the correct driver and licence.	39	113	159	61	12	2.72	0.95
	Demerit points are updated on a driver's record in a timely manner.	64	140	138	40	2	2.42	0.90
	A driver can easily verify or query their current point balance.	74	140	131	36	3	2.36	0.92
	Records held by the different agencies involved are consistent with one another.	42	136	142	57	7	2.61	0.93
	The system rarely loses, duplicates, or mismatches offence records.	48	132	139	55	10	2.60	0.97
Sanction processing	Points are assigned quickly after an offence is detected.	18	96	156	91	23	3.01	0.96
	Licence actions follow quickly once a driver reaches the point threshold.	45	123	131	75	10	2.69	1.00
	Sanctions are applied uniformly across comparable drivers and offences.	55	132	140	47	10	2.54	0.97
	The point thresholds that trigger action are predictable and clearly known.	28	120	131	90	15	2.85	0.99
	Sanctions are followed through to completion without being quietly dropped.	38	114	144	71	17	2.78	1.00
Acceptance	Enforcement officers treat drivers with respect during traffic stops.	18	74	155	99	38	3.17	1.00
	Enforcement decisions are made neutrally and without bias.	23	101	152	86	22	2.96	0.98
	The reasons for a sanction are explained clearly to the driver.	34	129	122	84	15	2.78	1.01
	Drivers have a fair opportunity to be heard or to contest an action.	22	81	145	113	23	3.09	0.98
	I trust that the enforcement system applies its rules fairly to everyone.	20	85	148	107	24	3.08	0.98
Deterrence	The demerit-point system has changed the way I drive.	24	91	149	97	23	3.01	0.99
	I believe that repeated offending will inevitably lead to real consequences.	42	104	155	66	17	2.77	1.00
	Exposure to enforcement has helped me correct risky driving habits.	51	126	136	59	12	2.62	1.00

Table 7. *Cont.*

Scale	Item	1	2	3	4	5	Mean	SD
	The system has reduced how often I take risks on the road.	22	95	155	89	23	2.99	0.97
	Overall, the enforcement chain from detection to sanction is credible and effective.	23	89	158	86	28	3.02	0.99

3.2. Diagnosis: Chain Index and Stability of the Ranking

Rescaled by Equation (1), the stage indices were 0.530 for detection, 0.385 for record integration and 0.445 for sanction processing, and the chain index of Equation (2) was 0.091 (95% bootstrap interval 0.082 to 0.100), far below any single stage (Figure 4). Adding acceptance as a fourth factor reduced it to 0.046. Sampling error in the stage means is small at this sample size, with intervals about 0.04 wide for each stage, and record integration was the lowest stage in every one of the 200,000 draws.

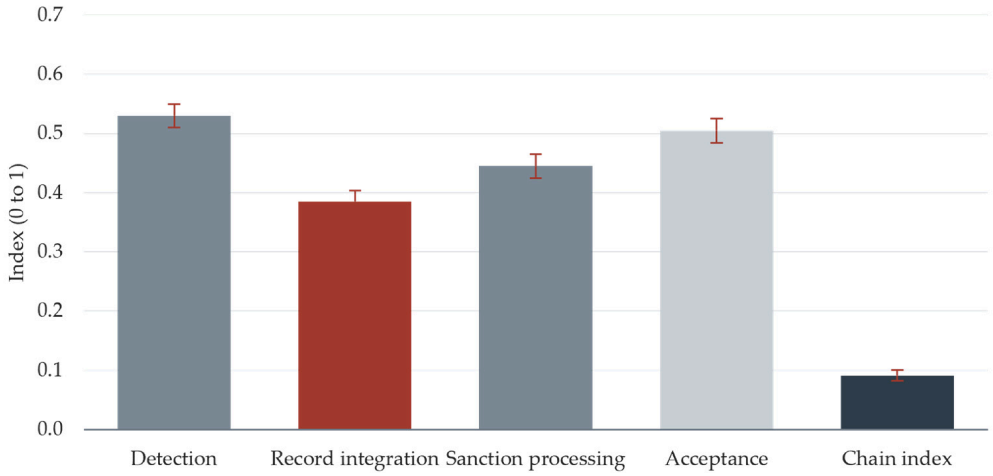

Figure 4. Stage indices and chain index with 95% intervals.

The ranking did not depend on the definition of the index (Table 8). Counting the share of respondents who agree gave a chain index of 0.012 and counting the share who do not disagree gave 0.228, against 0.091 for the rescaled mean, and record integration was the lowest stage under all three definitions; only the level of the chain index changed. The levels differ by an order of magnitude because the definitions place the neutral respondent differently, which is why the index compares stages and is not a rate of successful enforcement. The conclusion also depends on which items belong to which stage (Table 8). With the four overlapping stage items removed, record integration remained the lowest stage but the indices of record integration (0.411) and sanction processing (0.429) were close, and record integration was the lowest in 95% of bootstrap draws (94% to 96% across the family of item correlation matrices of Section 2.5). When the items on the timely posting of points and on checking the balance were reallocated to sanction processing and the item on recording the offence against the driver to record integration, sanction processing became the lowest (0.416 against 0.446 for record integration) in every draw. Removing the predictability item raised detection from 0.530 to 0.545 and changed neither the other stages nor the lowest stage. Record integration and sanction processing are therefore the weakest pair, and which of the two is lower depends on where the boundary between them is drawn.

Table 8. Stage indices and chain index under three definitions and three item checks.

Definition	Detection	Record Integration	Sanction Processing	Chain Index	Lowest Stage
Mean score rescaled to 0 to 1 (Equation (1))	0.530	0.385	0.445	0.091	Record integration
Share of respondents who agree	0.361	0.147	0.234	0.012	Record integration
Share of respondents who do not disagree	0.736	0.517	0.599	0.228	Record integration
Overlapping items removed	0.527	0.411	0.429	0.093	Record integration
Items reallocated to the stage each control addresses	0.527	0.446	0.416	0.098	Sanction processing
Predictability item removed	0.545	0.386	0.444	0.093	Record integration

The product treats the stages as independent. The three ratings were positively correlated (0.39 to 0.45), so the product is the lower end of a range: with the observed association the chain index lay between 0.091 and 0.385, and an illustrative Gaussian copula scenario placed it at 0.185, about 2.0 times the product (Figure 5). These are dependence scenarios based on correlations between ratings and are not estimates of dependence between the stages for the same offence. The figure also shows the chain index under the other two definitions.

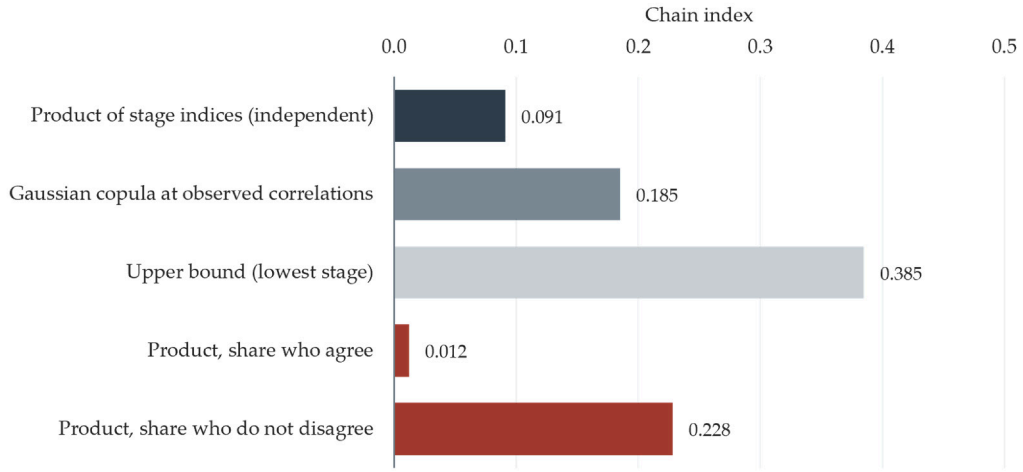

Figure 5. Chain index under dependence assumptions and three index definitions.

3.3. Diagnosis: Association with Perceived Deterrence

The four stage scores jointly explained 49.7% of the variance in perceived deterrence (R^2 = 0.497, shrinkage-corrected 0.485, $F(4, 379) = 93.5$), and variance inflation factors did not exceed 1.48, so collinearity did not distort the estimates. All four standardised coefficients were positive with intervals that excluded zero (Table 9). Sanction processing had the largest (0.29), followed by record integration (0.28), detection (0.19) and acceptance (0.19), but none of the six pairwise differences between coefficients exceeded two standard errors (the largest was 1.6), so the coefficients alone do not rank the stages. The intervals of the four coefficients overlap.

Table 9. Regression and dominance results, $n = 384$.

Stage	Standardised Coefficient (95% Interval)	Dominance Share, % (95% Interval)	Probability Ranked First
Detection	0.19 (0.11 to 0.27)	21 (14 to 29)	0.02
Record integration	0.28 (0.19 to 0.36)	29 (21 to 37)	0.29
Sanction processing	0.29 (0.21 to 0.38)	32 (24 to 41)	0.68
Acceptance	0.19 (0.11 to 0.27)	18 (12 to 26)	0.01

Table 9 also gives the dominance shares with intervals and the probability of ranking first from the resamples of the correlation matrix. Sanction processing accounted for 32% of the explained variance, record integration for 29%, detection for 21% and acceptance for 18%, and relative weights gave 32%, 29%, 21% and 19%. In the sample, sanction processing completely dominated the other three stages and record integration completely dominated detection and acceptance, whereas detection and acceptance did not dominate each other. Sanction processing exceeded record integration in 70% of resamples, so their order is uncertain, but the pair exceeded both other stages in 82%. The defensible reading is that sanction processing and record integration are the stronger correlates of perceived deterrence and detection and acceptance the weaker, without a firm order inside either pair.

The main associations above rest on the reported correlation matrix of the administered scales. As an exploratory sensitivity check, the analysis was repeated with the overlapping items removed. Its detail depends on how individual items correlate, which the available statistics do not fix (Section 2.5), so it is not a recomputation from respondent answers. With the overlapping items removed, the four stage scores explained a median of 45% of the variance in perceived deterrence across the family of item correlation matrices (range 40% to 50%; 44.7% for the common-correlation matrix, shrinkage-corrected 0.434). The dominance shares in the median matrix were 21% for detection, 27% for record integration, 33% for sanction processing and 19% for acceptance. Sanction processing and record integration both had larger shares than detection and acceptance in 89% of the matrices; this proportion fell from 98% to 75% as the item-level heterogeneity increased. Sanction processing ranked first in a median of 72% of resamples (range 14% to 97% across matrices) against 20% for record integration, so the order within the pair is not settled. With the items reallocated to the stage that the design assigns them to, sanction processing had a median dominance share of 39% (range 34% to 44%) and ranked first in a median of 98% of resamples (range 54% to 100%). Both checks keep sanction processing and record integration ahead of detection and acceptance in most item correlation structures but not in all of them. The pattern is therefore a likely one and not a confirmed one, because the respondent-level answers cannot be located and the check could not be computed from them. Because the same respondents rated every variable in one questionnaire, common-method variance cannot be excluded.

Figure 6 places each stage by its index and by its share of the explained variance. Record integration combines the lowest index with the second-largest share, sanction processing has a middle index and the largest share, and detection and acceptance have the highest indices and the smallest shares. The two diagnoses therefore name different stages, since record integration binds the chain and sanction processing is the largest contributor to deterrence, but they point to the same two stages for early repair.

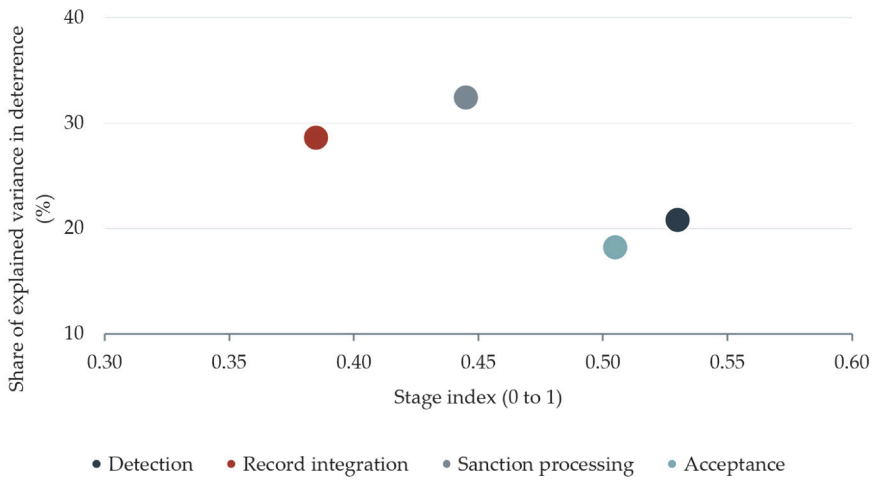

Figure 6. Stage index against share of explained variance in deterrence.

3.4. Diagnosis: Direct and Indirect Effects

Table 10 gives the path model of the chain. Record integration was predicted by detection (standardised coefficient 0.39, 95% interval 0.30 to 0.47), and sanction processing by both detection (0.30) and record integration (0.33). Detection and sanction processing stayed associated after record integration was held fixed (partial correlation 0.31, $p < 0.001$), so the chain is not strictly sequential: ratings of detection were linked to ratings of sanction processing directly, as well as through record integration, which is what would be expected if respondents hold a general view of how well the system works as well as a view of each stage. The model explained 15% of the variance of record integration and 28% of that of sanction processing.

Table 10. Standardised path coefficients of the chain.

Path	Coefficient	95% Interval
Detection to record integration	0.39	0.30 to 0.47
Detection to sanction processing	0.30	0.21 to 0.39
Record integration to sanction processing	0.33	0.24 to 0.42
Detection to deterrence (direct)	0.19	0.11 to 0.27
Record integration to deterrence (direct)	0.28	0.19 to 0.36
Sanction processing to deterrence (direct)	0.29	0.21 to 0.38
Acceptance to deterrence (direct)	0.19	0.11 to 0.27

Table 11 and Figure 7 split the effect of each stage on deterrence into a direct part and a part that passes through later stages. Detection had the largest total effect (0.43), and 55% of it passed through record integration and sanction processing. Record integration had a total effect of 0.37, of which 26% (interval 17 to 38%) passed through sanction processing. Sanction processing, the last stage, had only a direct effect (0.29). Because part of the effect of record integration passes through sanction processing, its total effect was larger than that of sanction processing although its direct coefficient was slightly smaller (0.28 against 0.29). A record integration rating that differed by the gap to the mean rating of detection was associated with a deterrence difference of 0.23 scale points (95% interval 0.18 to 0.28), against 0.10 points (0.07 to 0.13) for sanction processing, a ratio of 2.3 (interval 1.6 to 3.5). Detection, the highest-rated stage, had no gap to close. The gap to close was larger for record integration than for sanction processing, and the interval of the ratio takes the two gaps as

fixed and ignores the sampling error of the mean ratings. The path model therefore gives a third measure that favours record integration, but under an assumed process order and with overlapping intervals, whereas the regression alone does not separate the two stages. These direct, indirect and total effects describe associations among ratings collected at one time under an order taken from the process; they are not causal effects or expected benefits of intervention, and a smaller coefficient does not show that detection or fairness deserve less attention.

Table 11. Direct, indirect and total effects on deterrence.

Stage	Direct Effect	Indirect Effect	Total Effect (95% Interval)	Share Through Later Stages (%)	Associated Difference for the Gap to the Rating of Detection (95% Interval)
Detection	0.19	0.23	0.43 (0.34 to 0.50)	55	Not applicable (highest rating)
Record integration	0.28	0.10	0.37 (0.29 to 0.45)	26	0.23 (0.18 to 0.28)
Sanction processing	0.29	0.00	0.29 (0.21 to 0.38)	0	0.10 (0.07 to 0.13)

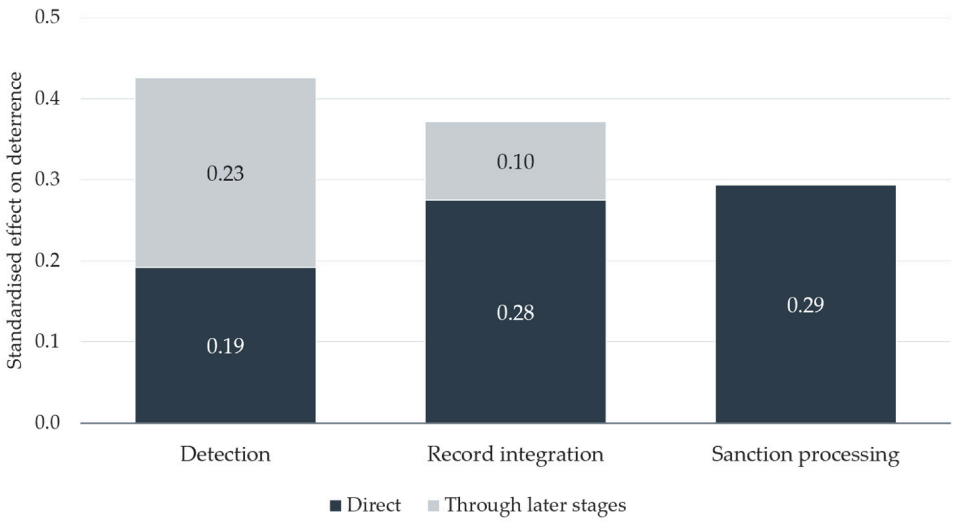

Figure 7. Direct effect and effect through later stages on deterrence.

3.5. Diagnosis: Repair Scenarios and Spending Rules

Table 12 and Figure 8 show what repair would do to the chain, with the stage indices in the order detection, record integration and sanction processing. A rise of 0.10 in the index of record integration lifted the chain index by 26% under independence, against 22% for sanction processing and 19% for detection, since the proportional gain is the increment divided by the current stage index (Equation (6)). Under the copula the same gains were 17%, 13% and 10%, so dependence reduces the gain from repair but leaves the order unchanged. Bringing record integration to the detection level lifted the chain index by 38% (copula: 24%), and bringing both weaker stages to that level lifted it by 64% (copula: 39%). Only a rating of 0.75 on every stage, which corresponds to a mean score of 4, moved the index to 0.42.

Table 13 sets two measures of repair side by side, which use different steps: the chain measure is per unit of index and the deterrence measure is per half a score point, equal to 0.125 of index. The chain measure is the sensitivity of the chain index to a stage. It was highest for record integration (0.236), then sanction processing (0.204) and detection

(0.171). The deterrence measure is the difference associated with a stage by the regression, which is an association and not a predicted effect. A half-point rise in the score of sanction processing or record integration corresponded to a perceived deterrence difference of 0.15 and 0.15 points, almost the same, and for detection and acceptance to 0.10 and 0.09. Because record integration lies furthest below the rating of 3.5, raising each stage to that rating favoured it most: the associated differences were 0.29 for record integration, 0.21 for sanction processing, 0.09 for acceptance and 0.07 for detection. A difference in that size in all four stages would be associated with a deterrence score of 3.54 against 2.88, which describes the sample and is not a forecast. Both measures place record integration and sanction processing ahead of detection. They do not show that detection or acceptance deserve less attention.

Table 12. Chain index under repair scenarios.

Scenario	Stage Indices	Chain Index, Independent	Chain Index, Copula	Gain, Independent (%)	Gain, Copula (%)
Current ratings	0.53/0.39/0.45	0.091	0.185	-	-
Record integration +0.10	0.53/0.49/0.45	0.114	0.216	26	17
Sanction processing +0.10	0.53/0.39/0.55	0.111	0.209	22	13
Detection +0.10	0.63/0.39/0.45	0.108	0.204	19	10
Record integration raised to the detection level	0.53/0.53/0.45	0.125	0.229	38	24
Sanction processing raised to the detection level	0.53/0.39/0.53	0.108	0.206	19	11
Both raised to the detection level	0.53/0.53/0.53	0.149	0.257	64	39
All three stages at 0.75	0.75/0.75/0.75	0.422	0.519	365	180

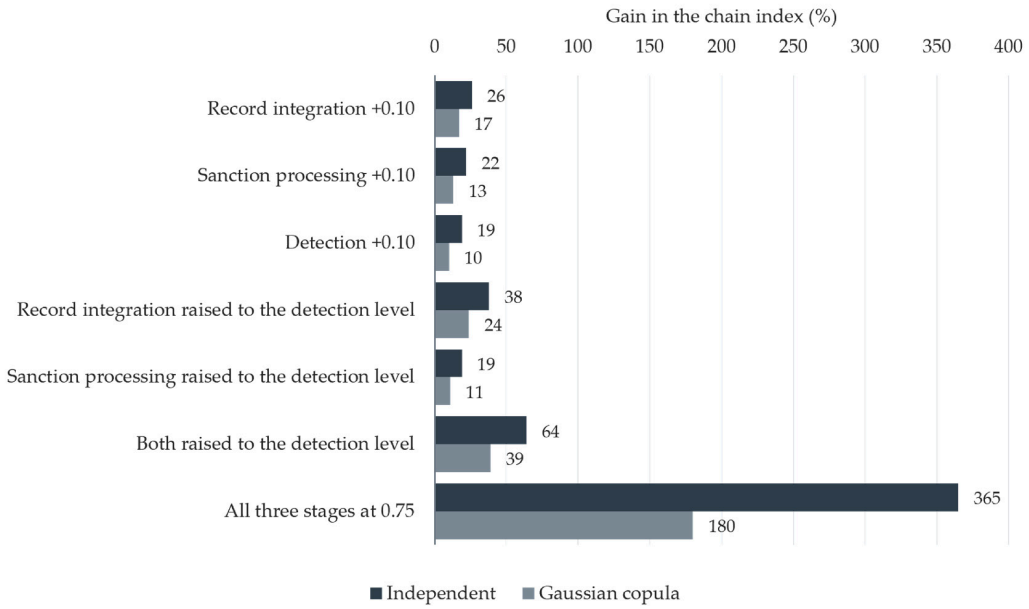

Figure 8. Gain in the chain index under seven repair scenarios.

Table 13. Two measures of repair for each stage.

Stage	Index	Chain Sensitivity	Chain Gain for +0.10	Relative Gain for +0.10 (%)	Associated Deterrence Difference for +0.5 (95% Interval)	Associated Difference to a Score of 3.5
Detection	0.53	0.171	0.0171	19	0.10 (0.05 to 0.14)	0.07
Record integration	0.39	0.236	0.0236	26	0.15 (0.10 to 0.19)	0.29
Sanction processing	0.45	0.204	0.0204	22	0.15 (0.10 to 0.19)	0.21
Acceptance	0.51	-	-	-	0.09 (0.05 to 0.13)	0.09

Table 14 and Figure 9 compare three rules for spending a budget of index points on the three technical stages. For a budget of 0.4, spending on the weakest stage first gave a chain index of 0.202, against 0.199 for an even split and 0.159 for concentrating on the strongest stage. The weakest-first rule exceeded the strongest-first rule by 27% but the even split by only 1.5%. Its gain over the strongest-first rule was 6% at a budget of 0.1 and 26% at 0.6; whereas, its gain over the even split stayed between 1.2% and 1.9% across the five budgets. The diagnosis therefore supports balanced repair led by the weakest stage and not a fine ordering.

Table 14. Chain index after spending a budget under three rules.

Budget (Index Points)	Weakest First	Even Split	Strongest First	Gain over Even Split (%)	Gain over Strongest First (%)
0.10	0.115	0.113	0.108	1.7	6
0.20	0.141	0.138	0.125	1.9	12
0.30	0.169	0.167	0.142	1.7	19
0.40	0.202	0.199	0.159	1.5	27
0.60	0.279	0.275	0.221	1.2	26

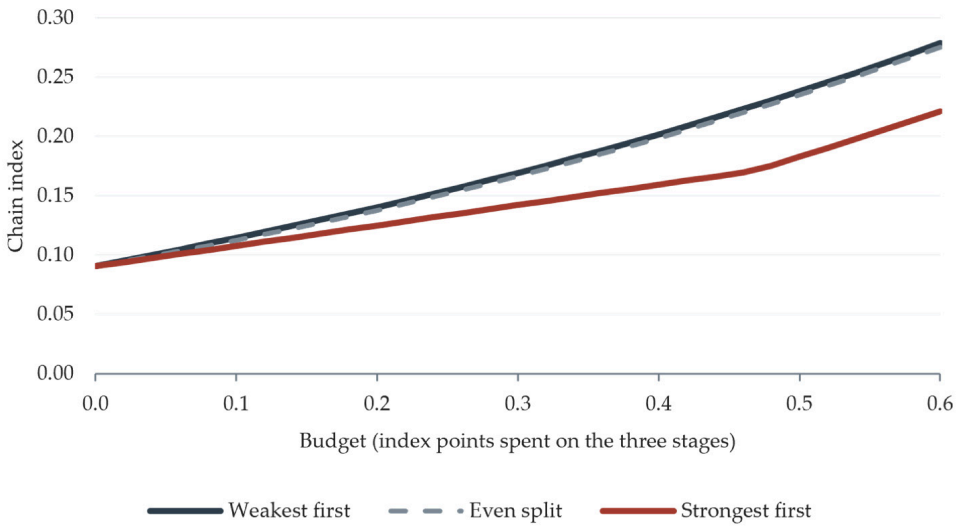

Figure 9. Chain index against budget under three spending rules.

3.6. Diagnosis: Lowest-Rated Statements and Officers' Accounts

Table 15 identifies where within the stages the ratings were lowest. Among the 20 stage statements, four of the six lowest-rated belong to record integration and two to sanction processing, and none belongs to detection or acceptance. The lowest was that a driver can easily verify a current point balance (mean 2.36, 95% interval 2.27 to 2.46; 56% disagreed), followed by timely updating of points (2.42) and uniform application of sanctions (2.54). The two lowest describe tasks that do not need new detection equipment, namely a way for drivers to see their record and a faster update of points after an offence.

Table 15. The six lowest-rated of the 20 stage statements.

Stage	Statement	Mean Rating	Disagreed (%)
Record integration	A driver can easily verify a current point balance	2.36 (2.27 to 2.46)	56
Record integration	Points are updated on the record in a timely manner	2.42 (2.33 to 2.51)	53
Sanction processing	Sanctions are applied uniformly across comparable cases	2.54 (2.45 to 2.64)	49
Record integration	The system rarely loses, duplicates or mismatches records	2.60 (2.51 to 2.70)	47
Record integration	Records held by different agencies are consistent	2.61 (2.52 to 2.71)	46
Sanction processing	Licence action follows quickly at the threshold	2.69 (2.59 to 2.79)	44

Table 16 sets the survey ratings beside the accounts of the six key informants (KI) along the chain. The accounts are explanatory and are not a validation of the indices: one or two informants spoke to each stage, and the accounts describe mechanisms that could produce the pattern of ratings and not their size. The informants described the failure of record integration in concrete terms: the offence and the licence sit in different systems, a driver can hold two entries, and a payment can clear while points remain. These are failures of linkage and reconciliation, of the kind that the lowest-rated items register. The accounts of delay in sanctioning are in the direction of the low rating of the speed of licence action. The accounts of detection describe presence without unpredictability, whereas the detection ratings were above the midpoint, and the accounts of acceptance describe partial legitimacy, which is consistent with a rating at the midpoint.

Table 16. Survey ratings and accounts of the six key informants by stage.

Stage	Survey Result	Account of Officers	Indicator
Detection	Mean 3.12; 36% agree	Enforcement concentrates on a few roads at known times, and offences pass unrecorded when devices are down or shifts are short (KI 3, KI 5).	Track unpredictability and device availability.
Record integration	Mean 2.54; 15% agree	Offence and licence sit in different systems, the same driver can have two entries, and a payment can clear while points remain (KI 1, KI 4).	Track the share of offences linked to the correct driver record.
Sanction processing	Mean 2.78; 23% agree	Action to suspend or call in a driver takes months, so the deterrent message is lost (KI 2).	Track time from offence to sanction.
Acceptance	Mean 3.02; 32% agree	A few bad encounters spread quickly and make drivers doubt that the system is fair to everyone (KI 6).	Explain reasons for each sanction and publish the route to contest it.

Table 17 sets out the seven failure modes that the informants described, with the survey statement that tests each. Five of the seven linked statements had means below the scale

midpoint of 3, and the three linked to record integration were the lowest, at 2.42, 2.60 and 2.61. Two statements were above the midpoint: that a detected offence is reliably recorded against the responsible driver (3.19) and that the system applies its rules fairly (3.08), so the accounts of offences passing unrecorded and of doubts about fairness were not supported by the ratings. Because six informants spoke to different stages, one or two to each, the match is illustrative and not a test, and the accounts serve to suggest which failure modes to examine in the record audit of Section 4.5.

Table 17. Failure modes described by key informants and the survey statement that tests each.

Failure Mode Described	Stage	Key Informants	Linked Survey Statement	Mean (Agreed, %)
Checks are concentrated on a few roads at known times	Detection	KI 3, KI 5	"Timing and location of checks are predictable enough to rely on"	2.91 (29)
Devices are down or shifts are short, so offences pass unrecorded	Detection	KI 3, KI 5	"A detected offence is reliably recorded against the responsible driver"	3.19 (36)
The offence and the licence are held in different systems	Record integration	KI 1, KI 4	"Records held by different agencies are consistent"	2.61 (17)
The same driver can hold two entries	Record integration	KI 1, KI 4	"The system rarely loses, duplicates or mismatches records"	2.60 (17)
A payment can clear while the points remain	Record integration	KI 1, KI 4	"Points are updated on the record in a timely manner"	2.42 (11)
Action to suspend or call in a driver takes months	Sanction processing	KI 2	"Licence action follows quickly at the threshold"	2.69 (22)
A few bad encounters spread and cast doubt on fairness	Acceptance	KI 6	"I trust the system applies its rules fairly"	3.08 (34)

3.7. Design: Effect of the Workflow on the Chain

Table 18 gives the stage and chain reliability under each package, starting from the baseline chain index of 0.091. Control points alone gave a median chain index of 0.30, 3.3 times the baseline, and backup routes alone 0.19, 2.1 times. Together they gave 0.40, with a 90% interval from 0.33 to 0.49, about 4.4 times the baseline (Figure 10). The combined gain, 0.312, is slightly larger than the sum of the two separate gains, 0.304, because stage reliabilities multiply across stages, even though within a stage a backup route adds less after the control points have acted. Control points lifted every stage in the model, whereas backup routes lifted only record integration and sanction processing, the two stages that have one, and left detection unchanged at 0.53. The two record integration elements alone (E2 and E3) gave an index of 0.17, 1.8 times the baseline, from a single stage.

Table 18. Chain index by package.

Package	Chain Index, Median	90% Interval	Detection	Record Integration	Sanction Processing
Baseline	0.09	0.09 to 0.10	0.53	0.39	0.45
Control points only	0.30	0.24 to 0.38	0.69	0.59	0.73
Redundancy only	0.19	0.16 to 0.22	0.53	0.55	0.64
Control points and redundancy	0.40	0.33 to 0.49	0.69	0.71	0.83
Record integration only (E2 and E3)	0.17	0.14 to 0.19	0.53	0.71	0.44

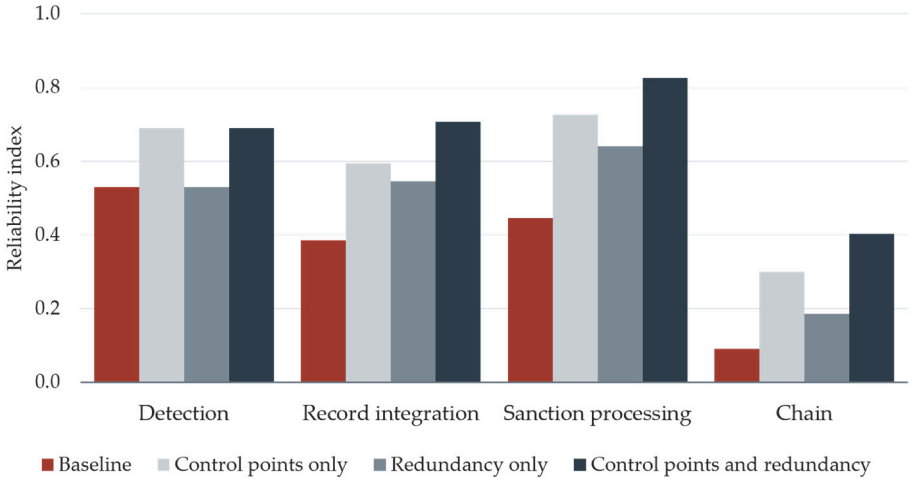

Figure 10. Stage and chain reliability by package, median values.

Table 19 and Figure 11 give the probability of reaching a target. The full workflow reached a chain index of 0.30 in more than 99% of draws, 0.40 in 53% and 0.50 in 3%. Control points alone reached 0.30 in 50% of draws, and backup routes alone reached 0.20 in 26% of draws and 0.30 in none. The baseline reached none of these targets in any draw. These probabilities are conditional on the assumed ranges and are not the probability that a real implementation would succeed.

Table 19. Probability (%) that the chain index reaches a target, by package.

Package	0.20 or More	0.30 or More	0.40 or More
Baseline	<1	<1	<1
Control points only	>99	50	2
Redundancy only	26	<1	<1
Control points and redundancy	>99	>99	53

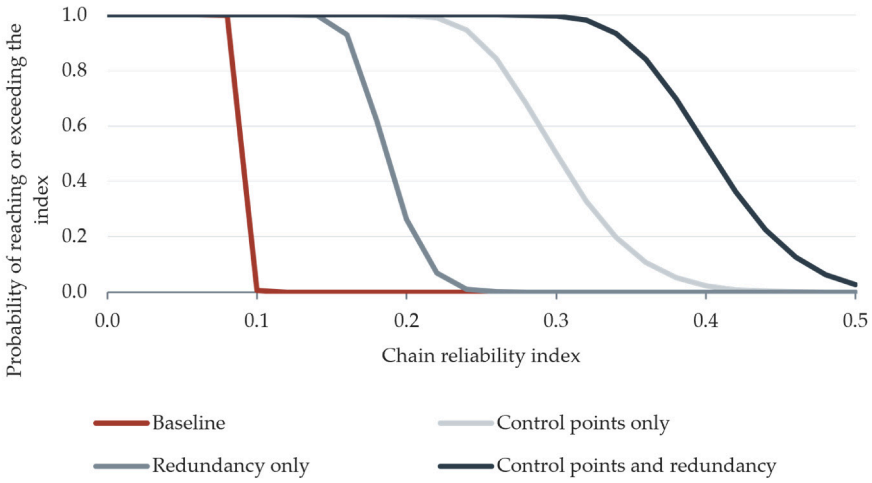

Figure 11. Probability that the chain index reaches a given value, by package.

3.8. Design: Gain of Each Element and Order of Introduction

Table 20 shows the gain of each element on its own. All six returned a positive gain whose 90% interval excludes zero. E2, the unique identifier, returned the largest median

gain, 0.049, followed by E5 at 0.040 and E4 at 0.038, and E1 and E6 the smallest, both 0.027. The intervals overlap, so the ranking is not secure under the assumed ranges: E2 has the highest median but cannot be separated clearly from E3, E4 or E5, and the lower gains of E1 and E6 overlap with all the others. E2 and E3 act on record integration, the stage with the lowest index, and each returned a higher ratio to baseline than E1, which acts on detection, as Proposition 1 states.

Table 20. Gain of each element on its own.

Element	Stage	Median Gain	90% Interval	Effort Tier
E1 Equipment monitoring	detection	0.027	0.016 to 0.043	2
E2 Unique identifier	record integration	0.049	0.029 to 0.077	3
E3 Second matching key	record integration	0.038	0.023 to 0.059	3
E4 Automatic posting	sanction processing	0.038	0.023 to 0.061	3
E5 Second notice channel	sanction processing	0.040	0.027 to 0.056	2
E6 Balance check	sanction processing	0.027	0.015 to 0.043	1

Table 21 and Figure 12 give two orders in which the elements could be introduced. Ordering by gain per unit of effort placed E6 first, because it is a procedural change, then E2, E1, E5, E3 and E4, and the chain index rose from 0.091 to 0.403 by the end. Ordering by gain alone gave E2, E5, E1, E4, E3 and E6, with the same end point. Neither order is better at every level of effort. The order by gain per unit of effort is ahead at 1, 2, 4, 6, 8, 9, 11 and 12 units of effort spent, the order by gain alone is ahead at 3, 5, 7, 10 and 13, and the two are equal at 14, because each step is a block of effort and the two orders complete their blocks at different points. The largest difference between them at any budget is 0.04 on the chain index, which is small against the total gain, so the choice depends on the budget and on how quickly a first gain is needed, and the order also depends on the effort weights, which are judgements (Section 3.12). The order is also not the baseline ranking of the elements. At baseline the gain per unit of effort ranks the elements as E6, E5, E2, E1, E4 and E3, but E5 moves from second to fourth place once E6, E2 and E1 have raised the other stages, as Propositions 1 and 4 together imply.

Table 21. Order of introduction by gain per unit of effort.

Step	Element	Stage	Effort Tier	Median Chain Index After the Step	Gain of the Step	Position by Gain Alone
1	E6 Balance check	sanction processing	1	0.118	0.027	6
2	E2 Unique identifier	record integration	3	0.183	0.065	1
3	E1 Equipment monitoring	detection	2	0.239	0.056	3
4	E5 Second notice channel	sanction processing	2	0.301	0.062	2
5	E3 Second matching key	record integration	3	0.357	0.056	5
6	E4 Automatic posting	sanction processing	3	0.403	0.046	4

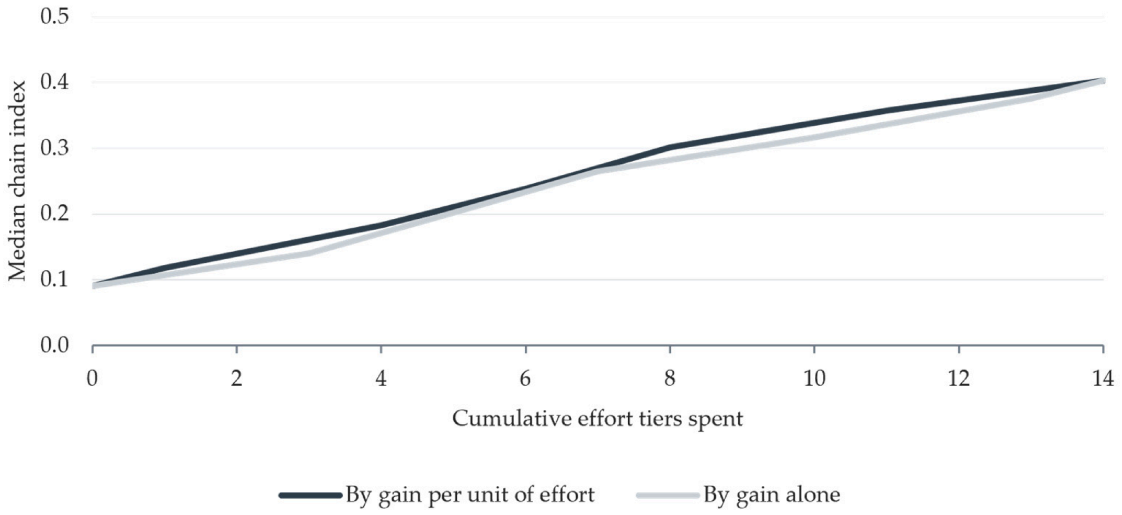

Figure 12. Median chain index against cumulative effort for two orders of introduction.

3.9. Design: Limits of Redundancy and of a Single Stage

The gain from a backup route fell in a straight line as the common-cause share rose (Figure 13). At the middle of the other ranges, E3 added 0.065 when the routes shared none of their failures, 0.033 when they shared half, and nothing when they shared all, and E5 fell from 0.062 to nothing. In the ranges assumed, which are 0.2 to 0.6 for E3 and 0.2 to 0.5 for E5, common cause removed between a fifth and three-fifths of the independent gain of E3 and between a fifth and a half of that of E5.

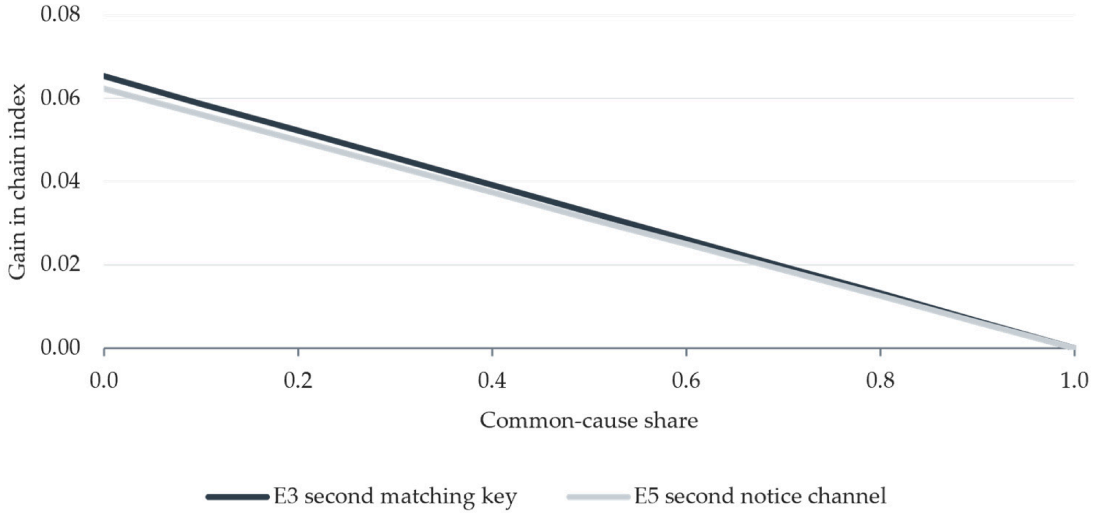

Figure 13. Gain from the two backup routes against common-cause share.

Table 22 shows what a single stage can do. Because the chain index cannot exceed the product of the other two stages, detection can reach at most 0.17, record integration 0.24 and sanction processing 0.20, if each were repaired alone and perfectly. A chain index of 0.15 would require the recovery of 41% of the failures of record integration, 52% of those of sanction processing or 74% of those of detection. A chain index of 0.20 is out of reach for detection and would require 75% of the failures of record integration or 96% of those of sanction processing, and 0.25 and above cannot be reached by any one stage. Repairing one stage is therefore not enough to lift the chain index much above 0.25 on this scale.

Table 22. Share of failures to recover to reach a target, by stage.

Target Chain Index	Detection	Record Integration	Sanction Processing
0.15	74	41	52
0.20	Not reachable	75	96
0.25	Not reachable	Not reachable	Not reachable
0.3	Not reachable	Not reachable	Not reachable
0.4	Not reachable	Not reachable	Not reachable
Highest index reachable	0.17	0.24	0.20

3.10. Design: Sensitivity and Check of the Propositions

Figure 14 ranks the parameters by their rank correlation with the chain index of the full workflow. The catch and correction rates of E1 and E2 were the most influential, with correlations of 0.48 and 0.34 for E1 and 0.45 and 0.32 for E2, followed by the reliability and common-cause share of E3, and the baseline ratings of the stages mattered least. Under the assumed ranges the result is therefore driven by what the control points on detection and record integration can catch and correct. The ranking reflects the widths of the ranges assumed, and the baseline ratings mattered little, partly because their standard errors are small. Rank correlation measures influence on the index and not on the choice of element, which is examined next.

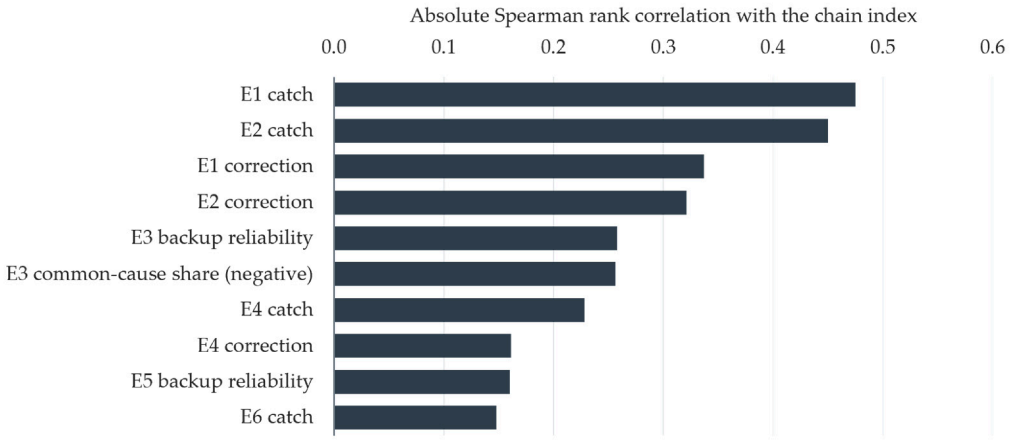

Figure 14. Rank correlation of the ten most influential parameters with the chain index.

Table 23 compares the gain that Equation (9) predicts, with every parameter at the middle of its range, with the median gain found by simulation. The predictions were within 3.5% of the simulated medians for all six elements and ranked them in the same order, so Proposition 1 describes the model well; the small excess of the predicted over the simulated gain is the effect of averaging over uniform ranges. Proposition 2 follows from the algebra of Equation (8), and Figure 13 illustrates it, as the gain at a common-cause share of 0.5 is half of that at zero. Proposition 3 is applied in Table 22. For Proposition 4, the three elements of sanction processing would raise the reliability of that stage by 0.529 if their separate gains at the middle of the ranges were added, but by 0.380 together, so that a later element on a stage adds less than it would alone. These agreements show that the algebra and the simulation of one model are consistent, which checks the implementation. They are not independent evidence for the propositions and do not test them against real enforcement data.

Table 23. Predicted and simulated gain of each element.

Element	Stage	Predicted Gain	Simulated Median Gain	Difference (%)
E1 Equipment monitoring	detection	0.0282	0.0274	+2.9
E2 Unique identifier	record integration	0.0508	0.0494	+2.8
E3 Second matching key	record integration	0.0392	0.0378	+3.5
E4 Automatic posting	sanction processing	0.0396	0.0385	+3.0
E5 Second notice channel	sanction processing	0.0405	0.0397	+2.0
E6 Balance check	sanction processing	0.0277	0.0269	+3.0

3.11. Design: Value of Information

The choice of the first element to introduce was treated as a decision, and Table 24 gives the share of the expected value of perfect information that each parameter carries, beside its rank correlation with the chain index. When the choice is made by gain alone, E2 is the best element in 51% of draws, and perfect information on every parameter would raise the expected gain of the choice by 12%. The catch rate of E2 carries 20% of that value, the catch rate of E4 6% and the correction rate of E2 5%. By contrast, the catch and correction rates of E1 had the largest rank correlations with the chain index of the full workflow (0.48 and 0.34, against 0.45 and 0.32 for E2), yet E1 carries almost none of the value (0%), because it is rarely the best first choice. When the choice is made by gain per unit of effort, E6 is the best element in 71% of draws, the value of information is small at 5% of the expected gain, and 29% of it lies in the catch rate of E6, the element that is cheapest to test in a pilot. The measurements with the greatest value for the decision are therefore the catch rates of E2 and E4 when the aim is the largest gain and the catch rate of E6 when effort is the constraint.

Table 24. Value of information by parameter.

Parameter	Rank Correlation with Chain Index	Share of Value of Information, Choice by Gain	Share of Value of Information, Choice by Gain per Effort
E2 unique identifier, catch rate	0.45	20	0
E4 automatic posting, catch rate	0.23	6	0
E2 unique identifier, correction rate	0.32	5	0
E3 second matching key, common-cause share	−0.26	1	0
E3 second matching key, backup reliability	0.26	1	0
E6 balance check, catch rate	0.15	0	29
E1 equipment monitoring, catch rate	0.48	0	0

3.12. Design: Sensitivity to the Assumptions

Table 25 gives the full workflow under the sixteen cases described in Section 2.11. In the thirteen alternative cases that keep the index definition of the diagnosis, the median chain index of the full workflow lay between 0.19 and 0.41 and the multiple of the baseline between 2.1 and 4.5. It was lowest, at 0.19, when the lower half of every range, a coverage of 30% to 70% of failures and a common-cause share of 0.5 to 0.9 applied together, and the target of 0.30 was then not reached in any draw. It was reached in 42% of draws with the

lower half of every range, in 31% with partial coverage and in 26% when E2 and E4 gave no benefit, against 100% as assumed. A chain index of 0.40 was reached in 53% of draws as assumed, in 22% under triangular distributions with the mode at the lower quarter, in 14% with the larger common-cause share and in none of the draws of the other less optimistic cases. The shape of the distributions and the correlation between parameters changed the median little (0.38 to 0.41) but, under correlation, widened the interval (0.29 to 0.56 at a correlation of 0.5), because parameters that move together move the chain together. Baselines taken from the alternative item allocations gave 0.41 in each case. Under the other two definitions of the index the full workflow reached 0.26 and 0.56, which reflects the different level of the baseline and not a different improvement. E2 had the highest median gain in all but three cases, and E5 had it when each control addressed only part of the failures of its stage, when E2 and E4 gave no benefit and when items were reallocated to the stage each control addresses, so the priority of the record integration controls is not independent of the assumptions. E6 was first by gain per unit of effort in every case but the first of these.

Table 25. Full workflow under sixteen cases.

Case	Baseline	Full Workflow (90% Interval)	Multiple	Reach 0.30 (%)	Reach 0.40 (%)	Top Gain	First by Effort
As assumed	0.091	0.40 (0.33 to 0.49)	4.4	100	53	E2	E6
Triangular, mode at middle	0.091	0.40 (0.35 to 0.46)	4.5	100	56	E2	E6
Triangular, mode at lower quarter	0.091	0.37 (0.33 to 0.43)	4.1	100	22	E2	E6
Correlated parameters (rho 0.5)	0.091	0.40 (0.29 to 0.56)	4.4	93	52	E2	E6
Correlated parameters (rho 0.8)	0.091	0.41 (0.28 to 0.59)	4.5	87	52	E2	E6
Lower half of every range	0.091	0.30 (0.27 to 0.33)	3.3	42	0	E2	E6
Recovery halved	0.091	0.23 (0.20 to 0.26)	2.5	0	0	E2	E6
Each control addresses 30 to 70% of stage failures	0.091	0.28 (0.24 to 0.34)	3.1	31	0	E5	E5
Common-cause share 0.5 to 0.9	0.091	0.35 (0.28 to 0.43)	3.8	88	14	E2	E6
Lower half, coverage 30 to 70% and common-cause 0.5 to 0.9	0.091	0.19 (0.16 to 0.22)	2.1	0	0	E2	E6
E2 and E4 give no benefit	0.091	0.28 (0.23 to 0.34)	3.1	26	0	E5	E6
Baseline overlap-free items	0.093	0.41 (0.34 to 0.49)	4.4	100	56	E2	E6
Baseline items reallocated to element stages	0.098	0.41 (0.35 to 0.49)	4.2	100	62	E5	E6
Baseline without predictability item	0.093	0.41 (0.34 to 0.49)	4.4	100	58	E2	E6
Baseline share who agree	0.012	0.26 (0.19 to 0.35)	21.0	22	1	E2	E6
Baseline share who do not disagree	0.228	0.56 (0.49 to 0.63)	2.4	100	100	E2	E6

Table 26 gives the order of introduction under four sets of effort weights for the three tiers and with equal weights. Every order reached the same final index of 0.40, so the weights change the route and not the destination. E6 came first and E4 last under every weighting that charges more for the higher tiers, and E2 moved from second place with weights of 1, 2 and 3 to third with 1, 2 and 4 and to fourth with 1, 2 and 6 and with 1, 3 and 9, as the price of its tier rose. With equal weights, which ranks by gain alone, E2 came first

and E6 last. The order of Section 3.8 is therefore a consequence of the assumed weights, and the effort weights are among the quantities to elicit (Section 4.5).

Table 26. Order of introduction under alternative effort weights.

Weights for Tiers 1, 2 and 3	Order of Introduction	Median Chain Index After Each Step	Total Effort
1, 2, 3 (as assumed)	E6, E2, E1, E5, E3, E4	0.12, 0.18, 0.24, 0.30, 0.36, 0.40	14
1, 2, 4	E6, E1, E2, E5, E3, E4	0.12, 0.15, 0.24, 0.30, 0.36, 0.40	17
1, 3, 9	E6, E1, E5, E2, E3, E4	0.12, 0.15, 0.19, 0.30, 0.36, 0.40	34
1, 2, 6	E6, E1, E5, E2, E3, E4	0.12, 0.15, 0.19, 0.30, 0.36, 0.40	23
1, 1, 1 (gain alone)	E2, E5, E1, E4, E3, E6	0.14, 0.20, 0.27, 0.32, 0.38, 0.40	6

4. Discussion

4.1. What the Diagnosis and the Design Show Together

The two parts answer different questions and point to the same two stages. Record integration had the lowest index in every resample, under every definition of the index and with the overlapping items removed, and it had the highest sensitivity of the chain index. Sanction processing had the largest share of explained variance in perceived deterrence, within sampling error of record integration. When the items were reallocated to the stage that the design assigns them to, sanction processing was the lowest. The two stages are therefore the weakest pair, and a diagnosis built on either alone would be incomplete. The modelled workflow concentrates on exactly these stages: E2 and E3 act on record integration, and E4, E5 and E6 on sanction processing. Under the assumed ranges, the six administrative controls moved the chain index in the model from 0.09 to about 0.4, a multiple of 4.4, without new detection equipment, and the multiple stayed between 2.1 and 4.5 in the thirteen alternative cases of Table 25 that keep the index definition. This is a statement about a constructed index and a model. It is not a forecast of enforcement performance and not a demonstrated improvement in deterrence or road safety. The full workflow reduced the shortfall of the chain index (one minus the index) from 0.909 to 0.597, a reduction of 34% on the index scale, unevenly: 34% for detection, 52% for record integration and 69% for sanction processing. An index of 0.40 is still low, and reaching 0.50 with equal stages would need 0.79 at each stage, so administrative measures of this kind would be necessary but not sufficient. The findings support treating record integration and sanction processing as candidate priorities for administrative audit and pilot evaluation. Detection and acceptance should be audited in the same exercise, because the survey associations and the illustrative simulations do not show that either deserves less attention. The order within the pair is not resolved by the data. That the two analyses agree on the pair while disagreeing on the first place is useful, since it suggests that the pair can be examined in parallel without waiting for a ranking that the data cannot give.

4.2. Records, Swiftness and Legitimacy in the Justice Chain

Record integration is the point at which detection becomes a consequence. A camera or officer that detects the offence adds little if the points are attached to the wrong record, to two records or to none, and a driver who cannot see a point balance cannot adjust behaviour to it. The lowest-rated items were exactly these, and the key informants described the same defects as ordinary operation and not as exceptions. The finding resembles the fragmentation of road traffic data across sources that each capture only part of the events [16], although that study concerned mortality data and not the linking of offences. A points system relies on a cumulative record, because the sanction is triggered by the sum of events and not by

any single one. Sanction processing carried the largest share of explained variance, and its weakest items were uniform treatment and the speed of licence action. Deterrence theory gives these the same status as certainty, since a sanction that is certain but slow loses force and one that is applied unevenly loses legitimacy [11]. The weakest item of acceptance was the explanation of reasons, which is the element of procedural fairness that a driver can observe directly [53].

4.3. Detection and the Allocation of Effort

Detection was rated highest, and its chain sensitivity of 0.17 was the lowest of the three technical stages. This does not show that detection is unimportant, since it is the input of the chain. At the ratings observed, the index suggests that further spending on cameras or patrols, while records and sanctions stay as they are, may add offences that the chain converts into consequences at a low rate. It also suggests that the consistency of detection, in particular equipment availability, may matter more than its extent. Neither the survey associations nor the simulations show that detection or procedural fairness should receive less attention, and both should be audited alongside the weaker stages. The gain modelled here is a gain in the certainty with which an offence reaches a consequence [11]; whether it changes driver behaviour and crash outcomes needs evaluation with outcome data. The lowest item, that a driver can verify a point balance, is the logic of E6, since a driver who cannot see a balance cannot adjust behaviour to it, and a statement of the offence, the points applied and the route to contest could be given with every sanction as an element of procedural fairness [53].

The officers' observation that coverage is concentrated at known places and times, and that offences pass when equipment is down, points to the consistency of detection, in particular device availability, and not its extent, which is the logic of E1. The survey item on the predictability of checks carried the same ambiguity: read as coverage it raises the detection rating, and read as an opening for evasion, as the officers described it, it lowers it, and removing it changed the detection index from 0.530 to 0.545 and gave no conclusion (Table 8). The index is a device for comparing stages and sequences of repair, and its level should not be read as a rate of enforcement. The lowest stage was the same under three definitions that differ in level by an order of magnitude, the product is the lower end of a range with an upper bound equal to the lowest stage, and under the copula the order of repair was unchanged although the gains were smaller. The comparison of repair rules adds a caution to the simple advice to repair the lowest-rated stage first. With equal costs the weakest-first rule exceeded the even split by only 1.5% at a budget of 0.4, because the best rule equalises the stages. The substantial gain is over concentrating on the strongest stage, which is a benchmark for how much allocation matters and not a description of how agencies spend. The advantage over concentrating on the strongest stage grew with the budget and levelled off near a budget of 0.4.

4.4. Elements Against Policy and Legislation

SI No. 53 of 2024 allocates points to categorised offences and provides for the suspension or cancellation of a licence when accumulated points pass a maximum [6,28]. The sanction is triggered by a balance, and a balance exists only if each offence was detected, linked to the right driver, adjudicated or settled as the law prescribes, and posted, which is the structure the chain index describes. On a planning scale where the chain index is 0.09, raising the points for an offence changes nothing for an offence that is not detected, linked or processed, which agrees with deterrence research that certainty carries more weight than severity [11]. The Data Protection Act of 2021 requires personal data to be accurate (Section 12(1)(d)) and gives a right to rectification (Section 59) [30], which agrees with the

logic of E2, E3 and E6 and with the low rating of record integration, although a perception of weakness does not show that the duty has been breached. Processing of personal data to detect or prosecute offences must be authorised by a written law and be necessary and proportionate (Section 40(1)) [30], so an additional identifier such as the national identity number in E3 cannot be adopted on reliability alone. The Electronic Communications and Transactions Act of 2021 supports E1, E4 and E5, because data messages are admissible with a weight that depends on the reliability of their generation and storage (Section 9) and the time of receipt of a record is set by the system designated by the addressee (Section 16) [31]. A second notice channel therefore needs a recorded contact point and time of receipt, and whether a text message is valid notice of an offence must be settled in enforcement procedures before E5 can be counted at the value assumed here. The regulator could also adopt as implementing guidance two indicators, the share of offences linked to the correct driver record and the time from offence to sanction, and report them to the lead agency against the targets of the Decade of Action [23,24].

4.5. Measurement Plan

The assumed ranges can be replaced by three measurements in this order, guided by the value of information. The first is structured expert elicitation [54]: officers working along the chain give low, most likely and high values for the catch rate, correction rate, backup reliability and common-cause share of each element, using a documented protocol with feedback between rounds, which gives expert ranges within weeks at little cost. The second is a small record audit: 200 to 300 offence records are followed through detection, record matching and sanction processing, which gives observed stage completion rates and, by testing a captured identifier and a second key against the licence and vehicle registers, the catch and correction rates of E2 and E3. The third is routine agency data, which would give stage counts at no extra cost if records could be linked along the chain. Engagement with the agency found its records fragmented across separate workflows and not linked along the chain, so routine data cannot yet supply stage counts and serve for now only to locate where the chain breaks and to choose the audit sample. The value of information in Table 24 shows which parameters to obtain first, and a repeat of the driver survey can test whether drivers' ratings follow the administrative indicators.

4.6. Implementation Framework

Figure 15 sets out a compact framework of five phases built on six design principles: ownership, so that each element has a responsible function and an indicator [55]; baseline measurement from the authority's own records in place of driver ratings; identity and data, so that the duties of accuracy and necessity are conditions of the design [30]; independence of backup routes, tested before their gain is counted; sequencing under a budget, with the order by gain per unit of effort as a starting point; and learning, through pilots that replace assumed ranges with measured ones. Its decision rules are that E3 is adopted only if its gain after common-cause loss exceeds its cost, that E5 is counted only if it does not share the contact record of the postal notice, and that work continues while the next element adds more per unit of effort than the alternative use of funds.

Table 27 sets out the same five phases with their activities, the finding that supports each, an indicator and a decision rule. The phases group the elements by stage, so their order differs from the order by gain per unit of effort, and they are reordered as measured values replace the assumed ranges.

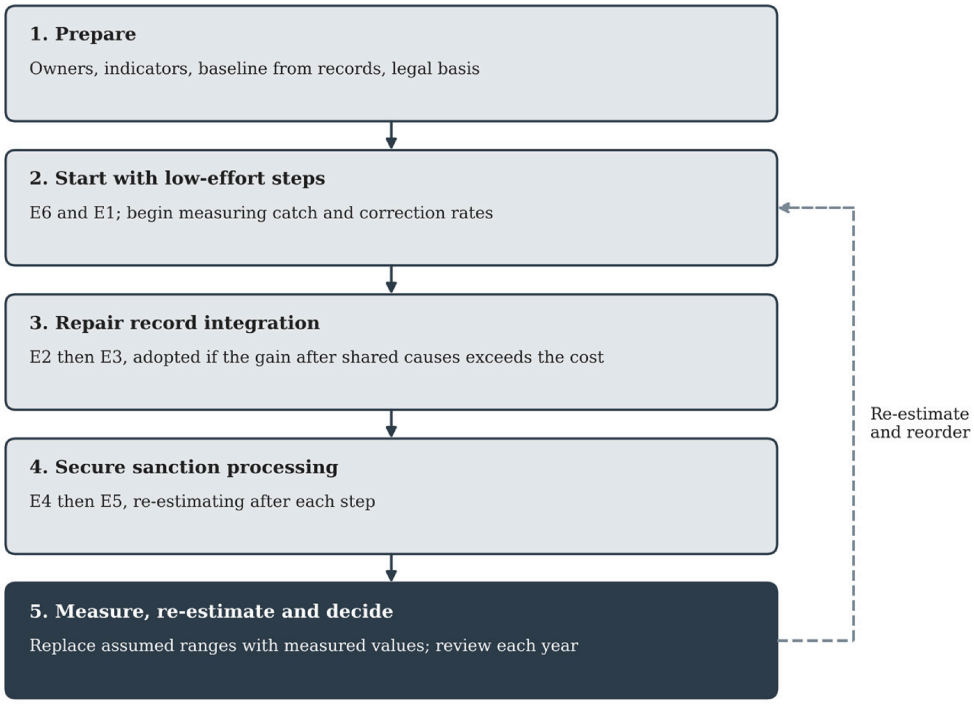

Figure 15. Implementation framework in five phases.

Table 27. Implementation framework: phases, activities and decision rules.

Phase	Activities	Supporting Finding	Indicator and Decision Rule
1. Prepare	Assign an owner to each element and publish the indicators of each element; measure the baseline of each stage from system records where they can be linked, and by a small record audit where they cannot; check the legal basis of each element	The chain index is set by the weakest stage and a single stage has a ceiling (Propositions 1 and 3)	Baseline indicator for each stage recorded. Decision rule: set priorities from the stage with the lowest measured index; the effect of added detection capacity was not modelled.
2. Start with low-effort steps	E6 driver check of the balance and E1 equipment monitoring; measure the catch rate of E6 in a pilot	E6 is first in the order by gain per unit of effort and its catch rate carries most of the value of information for that choice (Table 24)	Share of challenges resolved in the set period; share of device-hours in service. Decision rule: record the catch and correction rate of each control from the first month, and obtain expert ranges for E2 and E4 first.
3. Repair record integration	E2 unique identifier, then E3 second matching key	Record integration has the lowest index and E2 the highest single median gain (Table 20)	Share of offences linked to a verified driver record; share of unmatched offences resolved by the second key. Decision rule: adopt E3 only if its gain after common-cause loss exceeds its cost.

Table 27. *Cont.*

Phase	Activities	Supporting Finding	Indicator and Decision Rule
4. Secure sanction processing	E4 automatic posting, then E5 second notice channel, re-estimating after each step	Later elements on one stage add less than the first (Proposition 4); backup routes lose value to shared causes (Proposition 2)	Share of offences with points posted within one day; share of notices confirmed as received. Decision rule: confirm that E5 does not use the same contact record as the postal notice before counting its gain.
5. Measure, re-estimate and decide	Replace the assumed ranges with elicited and audited values, rerun the simulation, update the order and review each year	The result is conditional on assumed ranges (Table 4)	Chain index and stage indices each year. Decision rule: continue while the next element adds more per unit of effort than the alternative use of funds

4.7. Comparison with Related Work and Frameworks

The finding that enforcement fails at administrative stages is consistent with reviews of road safety management in Africa [2] and the multiplication of stage reliabilities gives quantitative form to the finding that certainty matters more than severity [11]. The worldwide meta-analysis concluded that points systems work to the extent that they are well administered [8]; the present results suggest which part of the administration is perceived as weakest and how stable that conclusion is. The fading of the effect in Spain [9] is what would be expected if record keeping and sanction speed erode, although a cross-sectional design cannot observe change, and the Italian study [10] estimated an effect at introduction, whereas this study estimates no behavioural effect and describes the chain that would produce one. The Safe System approach is deliberately broad [56], and the present framework can be read as an operational account of its enforcement layer, with a quantified order of work under assumed effects, while following the monitor-and-review cycle of ISO 39001 [55]. Legal frameworks assign duties, such as the service of documents on infringers in South Africa [5] and the exchange of offence information in the European Union [57], but do not specify how reliably each duty is performed; the same law can yield a substantially different chain index depending on how reliably the stages work. Unlike the outcome studies, this one rests on drivers' perceptions, which may differ from measured performance, and for this reason two administrative indicators are proposed.

4.8. Transferability and Implications for Practice

The workflow is defined by stages and functions. Wherever an offence must be detected, matched to a person and followed by a notice and a consequence, the same three stages and kinds of control apply, including camera programmes and points systems elsewhere. Road authorities in South Africa, Zambia and other countries where similar weaknesses are reported can replace the baseline indices and the assumed ranges with local values and rerun the simulation. The multiples and probabilities should not be transferred without that step, but the structure should. On these assumptions, a sequence for pilot evaluation under limited budgets could begin with procedural changes that cost little, such as the driver check of the balance, then link the offence record to a verified identifier, and add backup routes only where they do not depend on the same record, system or staff.

Three actions follow within an existing legal framework for police, licencing and adjudication bodies. The regulator could adopt, as implementing guidance, indicators for the time from offence to sanction and the share of offences linked to the correct driver record, and report both to the lead agency for road safety against the targets of the Decade of Action [23,24]. Reform of record integration would begin with a single driver identifier

shared by the detection, payment and licencing systems, a matching rule for duplicate entries, a reconciliation of payments with points, and a point balance that a driver can check on request, with lawful and accurate processing and a correction procedure for wrong records [30]. For transport and licencing administration, the owners and indicators in Table 2 give each control an accountable person and a measure computed from system records, which allows performance to be monitored as ISO 39001 expects [32]. For digital government, integration of records can be justified by the chain reliability that it adds and not only by convenience, and it should be tested first at the stage with the lowest index. For public investment, an agency that improves the conversion of detected offences into sanctions before buying more detection equipment avoids paying for offences that the chain cannot convert, which supports the efficient use of public funds towards the road death target of the Decade of Action [24]. The estimates should be read as a basis for pilots with measurement and not as forecasts. The actions listed are candidate priorities for administrative audit and pilot evaluation, and a change in the constructed index is to be distinguished from a demonstrated change in enforcement, deterrence or road safety, which only measured outcomes can show.

5. Limitations

The main limitation is that the effects of the controls are assumed. The ranges were set from failure mechanisms and checked against indirect evidence from other domains, and no measured effect of any of the six elements on a demerit point chain was available, so the multiples and probabilities are conditional and are not predictions. The cases of Table 25 show how far the results move when the ranges, their shape, their dependence and the share of failure modes addressed are changed, but they remain assumptions and not observations, and the model was not validated against administrative records, offence-processing rates or a pilot.

The baseline comes from drivers' perceptions, collected from one source at one time at high-throughput sites in one city during the roll-out of the system. The index is a comparison scale built from ratings and is not an observed rate of successful enforcement; a neutral rating, mapped to an index of 0.5, may reflect a lack of knowledge and not a middling performance, and 10.9% of respondents were unsure whether they had received points. The sample was not a probability sample, its stratum weights were planning allocations, 86.5% of respondents were men, and the survey intervals are indicative. The associations in the regression are not causal and may be inflated by common-method variance [45]; one of the five deterrence items asks whether the chain is credible, which overlaps in content with the predictors. The checks that removed or reallocated items are exploratory because the respondent-level answers cannot be located, and they relied on item counts, scale correlations and alpha, so their association results are reported as ranges over plausible item correlation structures (Section 3.3), and in a minority of those structures sanction processing and record integration did not lead detection and acceptance. With the overlapping items removed the deterrence mean moved from 2.88 to 2.87 and the chain index from 0.091 to 0.093, and the lowest stage depended on the allocation of items.

The model treats the stages as independent apart from the common-cause term, whereas the ratings were positively associated; those correlations are between perceptions and not between failures, and with a baseline nearer the upper bound of 0.39 the multiples would be smaller. Effort tiers are judgements and the order of introduction depends on them (Table 26). Routine agency records are fragmented, so stage counts could not be read from them, and the elicitation and the record audit proposed here have not been done. Drivers can observe some stages better than others, so the indices are perceived reliability. The repair comparison assumes equal cost per index point, which is a device

for ranking and not an estimate of cost, and the stage product assumes a multiplicative structure that the additive regression does not test. Only six key informants, recruited by snowball sampling, were interviewed, so their accounts help to explain the pattern of ratings and were not used to quantify it. The check of the propositions compares algebra with a simulation of the same model and cannot confirm them against real systems.

6. Conclusions

In Lusaka the perceived chain index of 0.09 was far below any single stage. Record integration was the lowest stage in every resample and under every definition of the index; with items reallocated to the stage each control addresses, sanction processing was the lowest, so the two form the weakest pair. Sanction processing and record integration were the stronger correlates of perceived deterrence, together accounting for 61% of the 50% of its variance that the four stages explained, and their order could not be separated. With the overlapping items removed, the pair led in most but not all plausible item correlation structures. Under assumed effects, a workflow of four control points and two backup routes gave a median chain index of 0.40 in the model, about 4.4 times the baseline, with 90% of draws between 0.33 and 0.49, and between 0.19 and 0.41 in thirteen alternative cases that keep the index definition. Within the model, the gain from an improvement is largest at the weakest stage, a backup route loses value in proportion to the failures it shares with the main route, no single stage can lift the chain index to 0.25 alone, and the order of introduction depends on the budget and on the effort weights. These results describe a constructed index and are not predictions of enforcement performance, deterrence or road safety. They identify candidate priorities for administrative audit and pilot evaluation: measuring the catch rates of the unique identifier and of automatic points posting, by structured expert elicitation and a small record audit, before costly systems are bought, and tracking the share of offences linked to the correct driver record and the time from offence to sanction. Detection and acceptance should be audited in the same exercise, because the data do not show that they deserve less attention. The approach applies to any enforcement chain whose stages can be rated or measured.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/vehicles8120317/s1, Supplementary File S1: Comprehensive research methodology and reproducibility documentation supporting the manuscript, *"Diagnosing and Raising Road Traffic Demerit Point Enforcement Reliability"*, including the research strategy, sampling procedures, complete questionnaire instruments, scoring protocols, key informant interview guides, analytical specifications, robustness and sensitivity assessments, simulation assumptions, and supporting Tables S1–S14.

Author Contributions: Conceptualization, K.A.L.; methodology, K.A.L.; software, K.A.L.; validation, K.A.L.; formal analysis, K.A.L.; investigation, K.A.L., M.P., E.N.K. and A.M.; resources, K.A.L.; data curation, K.A.L.; writing, original draft preparation, M.P., E.N.K. and A.M.; writing, review and editing, K.A.L., M.P., E.N.K. and A.M.; visualisation, K.A.L.; supervision, M.P.; project administration, K.A.L. and M.P.; funding acquisition, K.A.L. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was an undergraduate dissertation conducted as supervised academic research in the Department of Civil and Environmental Engineering, University of Zambia. At the School of Engineering, the topics of undergraduate dissertations receive an umbrella ethical approval from the School once the proposals have been approved, and the fieldwork was authorised under this arrangement through the Head of Department and the Academic Dean's Office. No separate approval or exemption reference number or date was issued for this study. Participants

were adult drivers and key informants, no sensitive personal data were collected, and questionnaires carried no direct identifiers. Consent is reported separately below.

Informed Consent Statement: Verbal informed consent was obtained from all drivers and key informants before participation. Questionnaires carried no direct identifiers, and informants are reported by non-identifying codes.

Data Availability Statement: The original contributions presented in this study are included in the Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank the drivers and the officers who took part in the study. The authors used Claude (Anthropic, Version 2.9939.2, build d3e504) to prepare the scripts that run the analyses and the simulation and draw the figures, to check the internal consistency of the reported numbers and to copy-edit the text. The authors are responsible for the content.

Conflicts of Interest: The authors declare no conflicts of interest. Furthermore, the authors K.A.L., M.P., E.N.K. and A.M. declare that funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CI	Confidence Interval
E1–E6	Enforcement Workflow Elements 1–6
EVPI	Expected Value of Perfect Information
ISO	International Organization for Standardization
KI	Key Informant
LMICs	Low-and Middle-Income Countries
RTSA	Road Transport and Safety Agency
SD	Standard Deviation
SE	Standard Error
SI	Statutory Instrument
VIF	Variance Inflation Factor

References

1. World Health Organization. *Global Status Report on Road Safety*; World Health Organization: Geneva, Switzerland, 2023.
2. Small, M. *Managing Road Safety in Africa: A Framework for National Lead Agencies*; Africa Transport Policy Program; SSATP: Washington, DC, USA, 2014.
3. Bonnet, E.; Lechat, L.; Ridde, V. What interventions are required to reduce road traffic injuries in Africa? A scoping review of the literature. *PLoS ONE* **2018**, *13*, e0208195. [CrossRef] [PubMed]
4. Staton, C.; Vissoci, J.; Gong, E.; Toomey, N.; Wafula, R.; Abdelgadir, J.; Zhou, Y.; Liu, C.; Pei, F.; Zick, B.; et al. Road traffic injury prevention initiatives: A systematic review and metasummary of effectiveness in low and middle income countries. *PLoS ONE* **2016**, *11*, e0144971. [CrossRef] [PubMed]
5. Khosa, D. The Future of South African Administration in the Context of the Implementation of the Administrative Adjudication of Road Traffic Offences. *J. Public Adm.* **2025**, *60*, 955–965. [CrossRef]
6. Government of Zambia. Road Traffic (Demerit Point System) Regulations, Statutory Instrument No. 53 of 2024. *Government Gazette*, 30 August 2024.
7. Capital, F.M. Motorists to Receive Traffic Fines via SMS, Email Under New NTSA Enforcement System. 2026. Available online: https://www.capitalfm.co.ke/business/2026/05/motorists-to-receive-traffic-fines-via-sms-email-under-new-ntsa-enforcement-system/ (accessed on 3 November 2026).
8. Castillo-Manzano, J.I.; Castro-Nuño, M. Driving licenses based on points systems: Efficient road safety strategy or latest fashion in global transport policy? A worldwide meta-analysis. *Transp. Policy* **2012**, *21*, 191–201. [CrossRef]
9. Castillo-Manzano, J.I.; Castro-Nuno, M.; Pedregal, D.J. An econometric analysis of the effects of the penalty points system driver's license in Spain. *Accid. Anal. Prev.* **2010**, *42*, 1310–1319. [CrossRef] [PubMed]
10. De Paola, M.; Scoppa, V.; Falcone, M. The deterrent effects of the penalty points system for driving offences: A regression discontinuity approach. *Empir. Econ.* **2013**, *45*, 965–985. [CrossRef]

11. Nagin, D.S. Deterrence in the twenty-first century. *Crime. Justice* **2013**, *42*, 199–263. [CrossRef] [PubMed]
12. Zaal, D. *Traffic Law Enforcement: A Review of the Literature*; Monash University Accident Research Centre: Clayton, Victoria, Australia, 1994.
13. Apel, R. Sanctions, perceptions, and crime: Implications for criminal deterrence. *J. Quant. Criminol.* **2013**, *29*, 67–101. [CrossRef]
14. Rausand, M.; Hoyland, A. *System Reliability Theory: Models, Statistical Methods, and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2003; Volume 396.
15. Reason, J. Human error: Models and management. *BMJ* **2000**, *320*, 768–770. [CrossRef] [PubMed]
16. Mwale, M.; Mwangilwa, K.; Kakoma, E.; Iaych, K. Estimation of the completeness of road traffic mortality data in Zambia using a three source capture recapture method. *Accid. Anal. Prev.* **2023**, *186*, 107048. [CrossRef] [PubMed]
17. Wilson, C.; Willis, C.; Hendrikz, J.K.; Le Brocque, R.; Bellamy, N. Speed cameras for the prevention of road traffic injuries and deaths. *Cochrane Database Syst. Rev.* **2010**, *11*, CD004607. [CrossRef] [PubMed]
18. Godthelp, H.; Ksentini, A. Specific road safety issues in low-and middle income countries (LMICs): An overview and some illustrative examples. *Traffic Saf. Res.* **2024**, *8*, e000068. [CrossRef]
19. Heydari, S.; Hickford, A.; McIlroy, R.; Turner, J.; Bachani, A.M. Road safety in low-income countries: State of knowledge and future directions. *Sustainability* **2019**, *11*, 6249. [CrossRef]
20. Bliss, T.; Breen, J. *Country Guidelines for the Conduct of Road Safety Management Capacity Reviews and the Specification of Lead Agency Reforms, Investment Strategies and Safe System Projects*; World Bank Global Road Safety Facility: Washington, DC, USA, 2009.
21. Flyvbjerg, B. Five misunderstandings about case-study research. *Sosiol. Tidsskr.* **2004**, *12*, 117–142. [CrossRef]
22. Yin, R.K. *Case Study Research and Applications*; SAGE: Thousand Oaks, CA, USA, 2018; Volume 6.
23. United Nations General Assembly. *Improving Global Road Safety*; Resolution A/RES/74/299; United Nations: New York, NY, USA, 2020. Available online: https://digitallibrary.un.org/record/3879711?v=pdf (accessed on 16 October 2026).
24. World Health Organization; United Nations Regional Commissions. *Global Plan for the Decade of Action for Road Safety 2021–2030*; World Health Organization: Geneva, Switzerland, 2021. Available online: https://www.who.int/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2021-2030/ (accessed on 16 October 2026).
25. African Union. *African Road Safety Charter*; African Union Commission: Addis Ababa, Ethiopia, 2016. Available online: https://au.int/en/treaties/road-safety-charter (accessed on 16 October 2026).
26. Government of Zambia. *Road Traffic (Amendment) Act, No. 8 of 2022*; Government Printer: Lusaka, Zambia, 2022. Available online: https://www.parliament.gov.zm/node/10311 (accessed on 16 October 2026).
27. Government of Zambia. Road Transport and Safety Agency Act No. 13 of 2002. 2002. Available online: https://zambialii.org/akn/zm/act/2002/13/eng@2002-12-31 (accessed on 16 October 2026).
28. Road Transport and Safety Agency. Public Notice: Demerit Point System. 2024. Available online: https://www.rtsa.org.zm/rtsa/about/ (accessed on 1 November 2026).
29. National Assembly of Zambia. Tuesday, 15th July, 2025: Sensitisation exercise by the Road Transport and Safety Agency. Parliamentary Debates. 2025. Available online: https://www.parliament.gov.zm/node/12493 (accessed on 3 November 2026).
30. Government of Zambia. Data Protection Act No. 3 of 2021. 2021. Available online: https://zambialii.org/akn/zm/act/2021/3/eng@2021-03-24 (accessed on 3 November 2026).
31. Government of Zambia. Electronic Communications and Transactions Act No. 4 of 2021. 2021. Available online: https://zambialii.org/akn/zm/act/2021/4/eng@2021-03-24 (accessed on 16 October 2026).
32. *ISO 39001:2012*; Road Traffic Safety (RTS) Management Systems: Requirements with Guidance for Use. International Organization for Standardization: Geneva, Switzerland, 2012. Available online: https://www.iso.org/standard/44958.html (accessed on 16 October 2026).
33. Cochran, W.G. *Sampling Techniques*, 3rd ed.; John Wiley and Sons: Hoboken, NJ, USA, 1977.
34. Cronbach, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika* **1951**, *16*, 297–334. [CrossRef]
35. Biernacki, P.; Waldorf, D. Snowball sampling: Problems and techniques of chain referral sampling. *Sociol. Methods Res.* **1981**, *10*, 141–163. [CrossRef]
36. Norman, G. Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* **2010**, *15*, 625–632. [CrossRef] [PubMed]
37. Carifio, J.; Perla, R. Resolving the 50-year debate around using and misusing Likert scales. *Med. Educ.* **2008**, *42*, 1150. [CrossRef] [PubMed]
38. Rönkkö, M.; Cho, E. An updated guideline for assessing discriminant validity. *Organ. Res. Methods* **2022**, *25*, 6–14. [CrossRef]
39. Esary, J.D.; Proschan, F.; Walkup, D.W. Association of random variables, with applications. *Ann. Math. Stat.* **1967**, *38*, 1466–1474. [CrossRef]
40. Nelsen, R.B. *An Introduction to Copulas*; Springer: Berlin/Heidelberg, Germany, 2006.
41. Yin, P.; Fan, X. Estimating R2 shrinkage in multiple regression: A comparison of different analytical methods. *J. Exp. Educ.* **2001**, *69*, 203–224. [CrossRef]

42. Budescu, D.V. Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychol. Bull.* **1993**, *114*, 542. [CrossRef]
43. Azen, R.; Budescu, D.V. The dominance analysis approach for comparing predictors in multiple regression. *Psychol. Methods* **2003**, *8*, 129. [CrossRef] [PubMed]
44. Johnson, J.W. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivar. Behav. Res.* **2000**, *35*, 1–19. [CrossRef] [PubMed]
45. Podsakoff, P.M.; MacKenzie, S.B.; Lee, J.Y.; Podsakoff, N.P. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J. Appl. Psychol.* **2003**, *88*, 879–903. [CrossRef] [PubMed]
46. Austin, K.P. *The Identification of Mistakes in Road Accident Records—Part Two*; ITS Working Paper No. 407; Institute for Transport Studies, University of Leeds: Leeds, UK, 1994. Available online: https://eprints.whiterose.ac.uk/2533/ (accessed on 16 October 2026).
47. Sorsavanh, T.; Liu, C.; Yamamoto, G.; Mori, Y.; Kobayashi, S.; Kuroda, T. Enhancing patient identification accuracy in shared child health records: A hybrid approach for the lao language context. *J. Med. Syst.* **2025**, *49*, 119. [CrossRef] [PubMed]
48. Bates, D.W.; Teich, J.M.; Lee, J.; Seger, D.; Kuperman, G.J.; Ma'Luf, N.; Boyle, D.; Leape, L. The impact of computerized physician order entry on medication error prevention. *J. Am. Med. Inform. Assoc.* **1999**, *6*, 313–321. [CrossRef] [PubMed]
49. Haynes, L.C.; Green, D.P.; Gallagher, R.; John, P.; Torgerson, D.J. Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *J. Policy Anal. Manag.* **2013**, *32*, 718–730. [CrossRef]
50. Mascagni, G.; Nell, C.; Monkam, N. *One Size Does Not Fit All: A Field Experiment on the Drivers of Tax Compliance and Delivery Methods in Rwanda*; ICTD Working Paper No. 58; Institute of Development Studies: Brighton, UK, 2017. [CrossRef]
51. Bell, S.K.; Delbanco, T.; Elmore, J.G.; Hong, P.G.A.; Ngo, L.H.; Urquhart, K.; McGeoch, E.; Bronsert, M.; Vogeli, C.; Walker, J.; et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. *JAMA Netw. Open* **2020**, *3*, e205867. [CrossRef] [PubMed]
52. Strong, M.; Oakley, J.E.; Brennan, A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: A nonparametric regression approach. *Med. Decis. Mak.* **2014**, *34*, 311–326. [CrossRef] [PubMed]
53. Tyler, T. Value-driven behavior. *Oxf. Handb. Law Econ.* **2017**, *1*, 402.
54. O'Hagan, A.; Buck, C.E.; Daneshkhah, A.; Eiser, J.R.; Garthwaite, P.H.; Jenkinson, D.J.; Oakley, J.E.; Rakow, T. *Uncertain Judgements: Eliciting Experts' Probabilities*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
55. Lie, A.; Tingvall, C. ISO 39001 Road Traffic Safety Management System, Performance Recording, and Reporting. In *The Vision Zero Handbook: Theory, Technology and Management for a Zero Casualty Policy*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 675–686.
56. Transport Research Centre, International Transport Forum, Organisation for Economic Co-operation and Development. *Towards Zero: Ambitious road Safety Targets and the Safe System Approach*; OECD Publishing: Paris, France, 2008.
57. European Parliament and Council of the European Union. Directive (EU) 2015/413 facilitating cross-border exchange of information on road-safety-related traffic offences. *Off. J. Eur. Union* **2015**, *L68*, 9–25.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.