The Arithmetic of Alert Fatigue: A Simulation of How Many Clinical Decision Support Rules a Clinician Can Usefully Carry

Dr. Priyadharshini P

Assistant Professor, Department of Community Medicine, KMCH Institute of Health Sciences and Research, Tamil Nadu, India.

Corresponding author: drpriya.cm@gmail.com

Received: 01 February 2026; Revised: 09 April 2026; Accepted: 16 April 2026; Published: 01 June2026

Abstract

Background: Clinical decision support rules are added to electronic health records one at a time, each justified on its own merits, and each imposing an alert burden on clinicians who are already overridden by most of what they see. Whether the aggregate remains useful is an arithmetic question that individual rule approval does not ask.

Objective: To quantify, by simulation, how alert volume, alert credibility and clinician engagement interact as rules accumulate, and to determine the conditions under which adding a rule reduces the clinical value the system delivers.

Methods: An agent-based simulation of 200 clinicians over 250 working days was constructed. Each clinician saw 20 patients daily; each deployed rule fired with defined sensitivity and specificity against a condition of defined prevalence. Engagement with an alert declined with the clinician's daily alert load and with the positive predictive value they had personally experienced, which was updated from the alerts they engaged with. Rule count, specificity and target prevalence were varied. Outcomes were alerts per clinician per day, alert positive predictive value, engagement rate, and true alerts actioned and missed per clinician per year.

Results: In the base configuration of ten rules at 95 % specificity against a 2 % prevalence condition, clinicians received 13.4 alerts per day with a positive predictive value of 26.8 %, engaged with 24.1 % of them, and actioned 216 true alerts per year while missing 683 and acting on 590 false ones. Specificity dominated every other lever: raising it from 95 % to 99.5 % cut the daily load from 13.4 to 4.6 alerts, raised positive predictive value from 27.0 % to 78.7 %, raised engagement from 24.3 % to 58.2 %, and more than doubled true alerts actioned from 219 to 524 per clinician per year. In the base model, adding rules never reduced the absolute number of true alerts actioned; an interior optimum appeared only when rules were allowed to differ in clinical value, with the best rules deployed first. Under that assumption the optimal portfolio was 5 rules at 90 % or 95 % specificity, 8 at 98 %, and 12 at 99.5 %.

Conclusions: Specificity, not rule count, is the governing parameter, and higher specificity is what earns the right to deploy more rules. Governance that approves rules individually against a clinical-merit test, without a portfolio budget, will reliably overshoot.

Keywords: Clinical Decision Support, Alert Fatigue, Alert Override, Simulation, Health Informatics, Electronic Health Records.

1. INTRODUCTION

Clinical decision support has an accumulation problem. A rule is proposed because a specific harm occurred or might occur, it is evaluated on whether it would catch that harm, and it is approved. The next rule is proposed on the same basis. Nothing in that process asks what happens to the previous rules when the new one is added, and the answer is that they all become slightly less effective, because the clinician's attention is a fixed quantity being divided among a growing number of claims on it.

The observable consequence is override. Reported override rates for medication alerts commonly exceed ninety per cent, and clinicians dismiss alerts faster than they could plausibly have read them. This is

frequently framed as a behavioural failing to be addressed by training or by making alerts harder to dismiss. It is more usefully understood as a rational response to a low base rate: a clinician who has learned that roughly one alert in four is worth acting on, and who receives a dozen a day, is behaving sensibly in triaging them quickly.

If that framing is right, then alert fatigue is a property of the alerting portfolio rather than of the clinician, and it is computable. The quantities involved — how many rules, how specific each is, how common the target condition is, how many patients are seen — are all known to the organisation deploying them, and the interaction between them determines whether the system as a whole delivers more clinical value than a smaller one would.

This study builds that computation. Three questions are asked. How do alert volume, credibility and engagement move as rules accumulate? Which lever — rule count, specificity, or target selection — most affects the value delivered? And is there a point at which adding a rule makes the system worse, not merely less efficient?

2. METHODS

2.1 Structure

An agent-based simulation was implemented in Python. Two hundred clinicians were simulated over 250 working days, each seeing 20 patients per day. Each deployed rule was evaluated against every patient, generating a true-positive alert when the condition was present and the rule detected it, and a false-positive alert when the condition was absent and the rule fired anyway. Rules were treated as independent, with identical sensitivity and specificity, targeting conditions of identical prevalence; heterogeneity is introduced in Section 3.5.

Table 1. Simulation parameters

Parameter	Base value	Range explored	Rationale
Clinicians	200	—	Sufficient for stable estimates
Working days	250	—	One clinical year
Patients per clinician per day	20	—	Representative outpatient load
Rules deployed	10	1 to 50	From a minimal set to a mature, accreted portfolio
Rule sensitivity	0.90	—	Rules are usually built to catch the target
Rule specificity	0.95	0.90 to 0.999	The parameter organisations rarely measure
Target prevalence	0.02	0.005 to 0.10	From rare safety events to common prescribing situations
Maximum engagement probability	0.95	—	Engagement with an alert seen in isolation
Load at which engagement halves	12 alerts/day	—	Calibrated to produce override rates in the reported range
Experienced-PPV weight	2.2	—	Strength of the credibility effect on engagement

2.2 The Engagement Model

The behavioural component is the only part of the model that is not simple accounting, and it encodes two mechanisms for which there is empirical support: clinicians engage less when they receive more alerts, and they engage less with a class of alert that has proved unreliable. Engagement probability was modelled as

$$P(\text{engage}) = P_{\max} \times [1 / (1 + load/L_{½})] \times \text{PPV}_{\text{experienced}}{}^{1/w} \quad (1)$$

where *load* is the clinician's alert count that day, $L_{½}$ the load at which engagement halves, and $\text{PPV}_{\text{experienced}}$ the positive predictive value that clinician has personally observed among the alerts they engaged with, updated after every day from a weak prior. The second term is what makes the model more than a queueing

calculation: credibility is learned from experience, so a portfolio that cries wolf trains clinicians to ignore it, including the parts of it that are reliable.

The functional form is a modelling choice rather than an estimated relationship, and the parameters were set so that the base configuration produces override behaviour in the range reported in the literature rather than fitted to any dataset. Section 4.3 states what this means for the interpretation of the results.

3. RESULTS

3.1 The Base Configuration

Table 2. Base configuration: 10 rules, 95 % specificity, 2 % prevalence

Quantity	Value
Alerts per clinician per day	13.4
Alert positive predictive value	26.8 %
Engagement rate	24.1 %
True alerts actioned per clinician per year	216
True alerts missed per clinician per year	683
False alerts acted upon per clinician per year	590

The base configuration reproduces the pattern reported from real deployments: a double-digit daily alert load, a positive predictive value around one in four, and an override rate around three in four. It also shows what that costs. For every true alert a clinician acts upon, roughly three are missed and roughly three false ones consume attention. The system is delivering 216 useful interventions per clinician per year while generating 3,350 alerts to do it.

3.2 Adding Rules

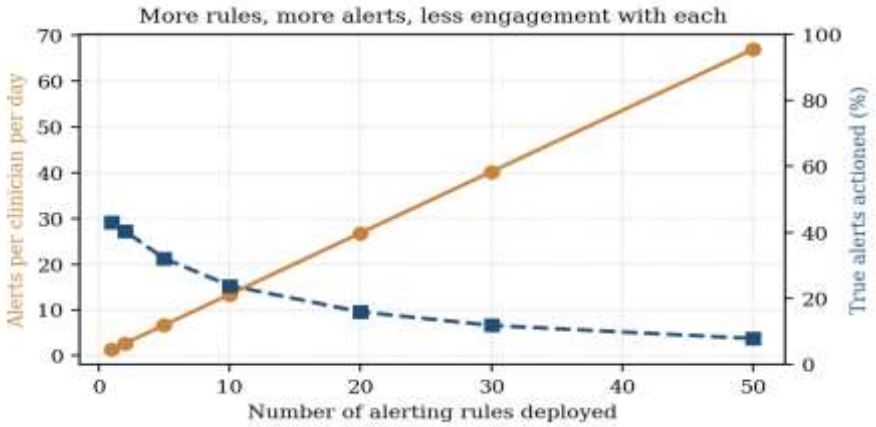

Figure 1. Daily alert load and the proportion of true alerts actioned, as rules accumulate.

Table 3. Effect of rule count (95 % specificity, 2 % prevalence)

Rules	Alerts/day	PPV (%)	Engagement (%)	True actioned/yr	True missed/yr	False actioned/yr
1	1.4	26.9	43.4	39	51	107
2	2.7	26.9	40.2	73	108	198
5	6.7	26.7	32.1	144	304	394
10	13.4	26.9	24.1	216	687	591
20	26.8	26.9	16.0	288	1,514	783
30	40.2	26.9	11.9	320	2,380	877
50	67.0	26.8	7.9	355	4,140	969

Three columns tell the story and they do not agree with one another. Alert volume grows linearly with rule count, as it must. Engagement collapses, from 43.4 % with a single rule to 7.9 % with fifty, because load rises and experienced credibility does not. And the absolute number of true alerts actioned nevertheless keeps rising, from 39 to 355.

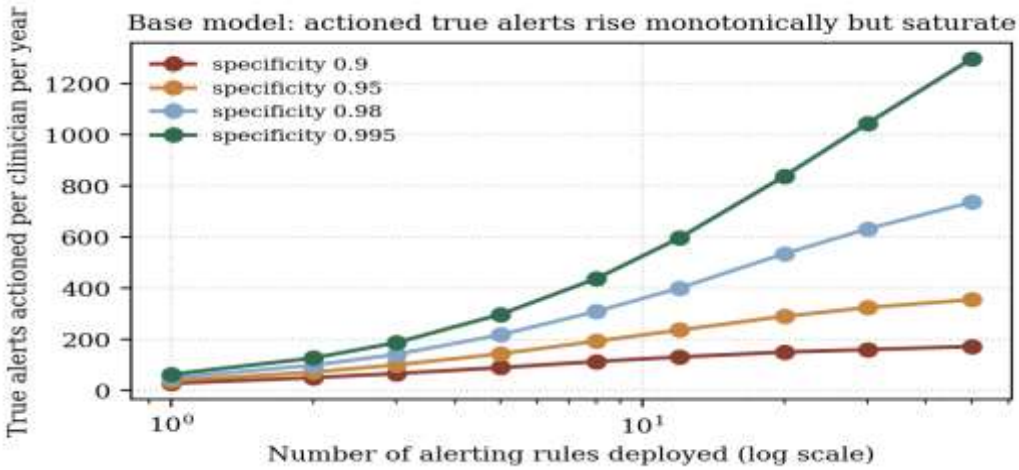

Figure 2. True alerts actioned per clinician per year against rule count in the base model, at four specificities.

This last result was not what the study set out to find, and it is reported as it came out. In a model where every rule is equally valuable, adding rules always increases the absolute number of true alerts actioned, because each new rule brings new true cases and the engagement penalty is sublinear. There is no interior optimum: on this criterion alone, fifty rules beat ten.

What the same table shows is what that gain costs. Going from ten rules to fifty increases actioned true alerts by 64 %, from 216 to 355, while increasing missed true alerts sixfold from 687 to 4,140 and false alerts acted upon by 64 % from 591 to 969. The system catches more in absolute terms and becomes far more wasteful per unit caught. Whether that trade is acceptable depends on a judgement the simulation cannot make; what it establishes is that "actioned true alerts" alone is the wrong criterion for portfolio decisions, because it is monotone in rule count and therefore cannot identify a stopping point.

3.3 Specificity dominates

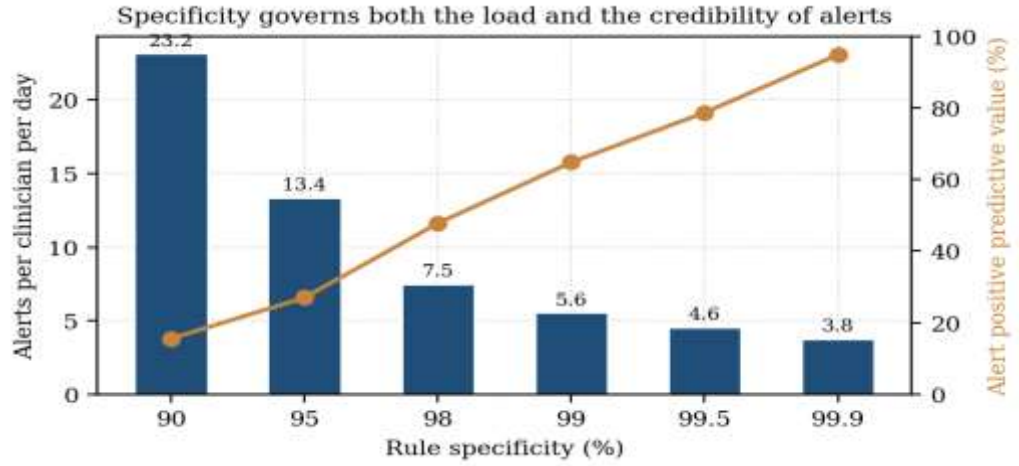

Figure 3. Daily alert load (bars) and alert positive predictive value (line) across rule specificities.

Table 4. Effect of specificity (10 rules, 2 % prevalence)

Specificity	Alerts/day	PPV (%)	Engagement (%)	True actioned/yr	True missed/yr	False actioned/yr
90.0 %	23.2	15.6	13.8	124	779	675
95.0 %	13.4	27.0	24.3	219	684	593
98.0 %	7.5	47.8	40.2	361	537	395
99.0 %	5.6	64.8	50.6	456	447	249
99.5 %	4.6	78.7	58.2	524	376	142
99.9 %	3.8	94.8	66.0	595	306	32

This is the finding with the clearest operational implication. Specificity moves every column at once and in the same direction. Raising it from 95 % to 99.5 % — a change invisible in any conventional description of a rule, which would be reported as "accurate" in both cases — cuts the alert load by two thirds, triples the positive predictive value, more than doubles engagement, more than doubles true alerts actioned, and cuts false alerts acted upon by three quarters.

Note what is happening mechanically, because it is why specificity is so much more powerful than the other levers. Sensitivity determines how many true alerts exist; specificity determines how many false ones accompany them. Because the target conditions are rare, false positives outnumber true ones by a large factor, so the alert load is almost entirely a function of specificity. And because credibility is learned from the mix, specificity also determines whether clinicians believe the system — which means it affects the true alerts twice, once by not burying them and once by making them believable.

The practical corollary is uncomfortable for how rules are usually built. A rule developer optimising sensitivity, which is the natural instinct when the rule exists to prevent a specific harm, is optimising the parameter that matters least and neglecting the one that determines whether the rule will be read.

3.4 Prevalence of the Target

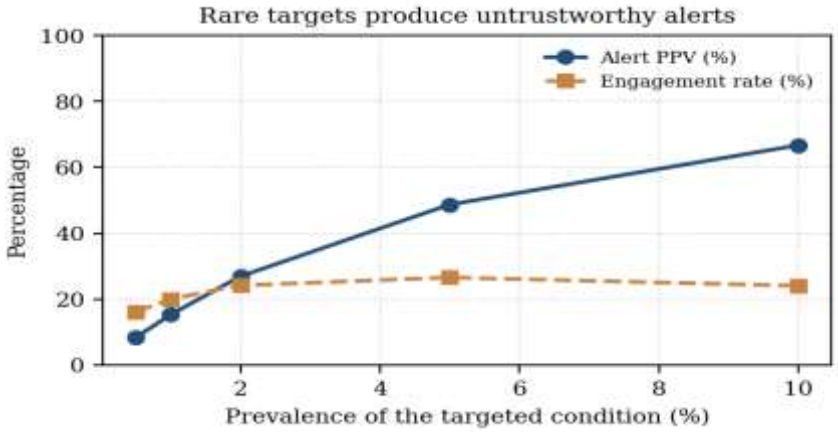

Figure 4. Alert positive predictive value and engagement rate across target prevalences.

Table 5. Effect of target prevalence (10 rules, 95 % specificity)

Prevalence	Alerts/day	PPV (%)	Engagement (%)	True actioned/yr	True missed/yr
0.5 %	10.9	8.4	16.1	37	190
1 %	11.7	15.3	20.1	90	358
2 %	13.4	26.9	24.2	218	684
5 %	18.5	48.7	26.6	599	1,652
10 %	27.0	66.7	24.1	1,084	3,418

Rules aimed at rare events are structurally weak, and not because they are poorly built. At a prevalence of 0.5 %, a rule with 90 % sensitivity and 95 % specificity produces alerts with a positive predictive value of

8.4 %: eleven alerts in twelve are false, clinicians learn this quickly, engagement falls to 16.1 %, and the rule delivers 37 actioned true alerts per clinician per year while generating roughly 2,700 alerts.

This is the alerting analogue of a familiar result in screening, and it has the same remedy. A rule targeting a rare event needs specificity far above the level that would be adequate for a common one, and if that specificity is unattainable the rule should probably not be an interruptive alert at all — a passive indicator, a report, or a batch review would deliver the same detection without spending the portfolio's credibility.

3.5 When Does Adding A Rule Make Things Worse?

Section 3.2 found no interior optimum under the assumption that all rules are equally valuable. That assumption is unrealistic in a specific and consequential way: organisations do not deploy rules in random order. The first rules addressed are the ones with the clearest clinical case, and later additions are progressively more marginal.

To test whether that changes the conclusion, the analysis was repeated with rules deployed best-first and the j-th rule assigned a clinical value of $1/j$. The value delivered by a portfolio of k rules is then the actioned true alerts per rule multiplied by the sum of the values, and because engagement falls as k rises while marginal value also falls, an interior optimum can exist. Table 6 and Figure 5 report it.

Table 6. Value-weighted true alerts actioned per clinician per year, rules deployed best-first

Rules	Spec. 90 %	Spec. 95 %	Spec. 98 %	Spec. 99.5 %
1	30.1	38.7	51.8	64.4
2	38.3	53.7	74.1	95.8
3	41.7	62.3	87.3	114.1
5	42.1	65.9	100.5	136.0
8	39.0	65.8	105.8	149.5
12	34.0	61.2	104.5	154.6
20	27.1	51.7	95.8	150.1
30	21.3	42.7	83.5	138.9
50	15.3	32.1	65.8	116.5
Optimal portfolio	5 rules	5 rules	8 rules	12 rules

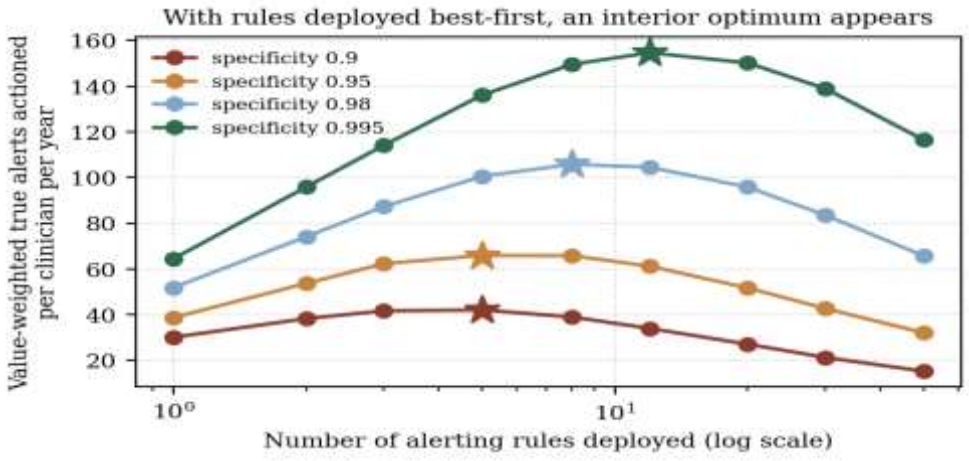

Figure 5. Value-weighted actioned true alerts against rule count, with the optimum starred, under best-first deployment.

An interior optimum appears at every specificity, and its location is the result worth taking away: 5 rules at 90 % and 95 % specificity, 8 at 98 %, and 12 at 99.5 %. Specificity is what buys the right to deploy more rules. An organisation running thirty rules at 95 % specificity is delivering about two thirds of the value it would deliver with five, and less than a third of what twelve rules at 99.5 % would deliver.

The conditional nature of this finding should be stated plainly rather than buried. The interior optimum exists because rules were assumed to differ in value and to be deployed in descending order of it. That assumption is realistic but it is an assumption, and it is doing the work: without it, Section 3.2 found that more rules are always better on the absolute criterion. What the two sections jointly establish is the necessary condition for an alerting paradox, which is more useful than a bare demonstration of one would have been.

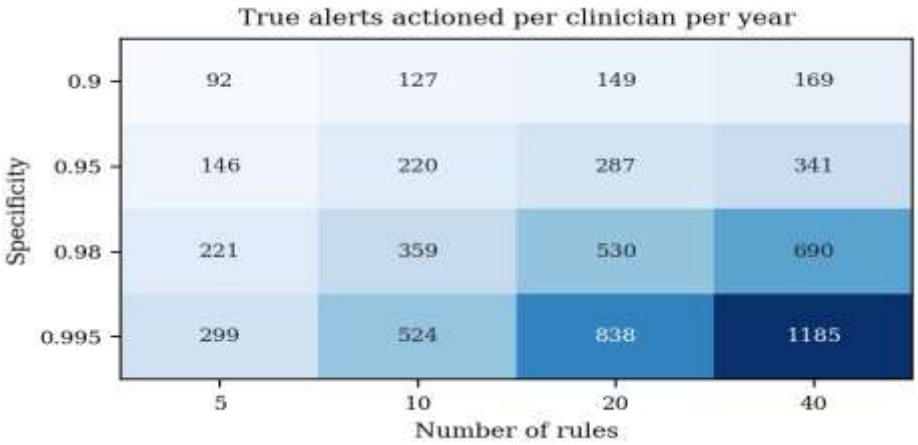

Figure 6. True alerts actioned per clinician per year across combinations of rule count and specificity.

4. DISCUSSION

4.1 Principal Findings

Alert fatigue in this model is not a behavioural deficiency but an equilibrium. Clinicians receiving thirteen alerts a day of which one in four is actionable engage with about a quarter of them, which is close to what deployed systems report, and the behaviour is a reasonable allocation of attention given the base rate they have experienced. Changing it by exhortation, or by making alerts harder to dismiss, addresses the symptom while leaving the base rate that produced it untouched.

Specificity is the dominant lever by a wide margin, and it acts twice: it determines the volume of alerts and it determines their credibility, which in this model is learned and therefore persistent. Sensitivity, which rule developers naturally optimise, determines only how many true alerts exist to be buried.

And rule portfolios have an optimal size that depends on their specificity, provided rules differ in value. At 95 % specificity that optimum was five rules; at 99.5 % it was twelve. Portfolios in deployed systems are frequently an order of magnitude larger.

4.2 Implications for Governance

Table 7. Implications

Finding	Implication	Governance action
Specificity moves every outcome at once	It is the parameter to measure and to specify	Require a measured specificity before approval; refuse rules that cannot report one
Rules aimed at rare events have low PPV regardless of quality	Interruptive alerting is the wrong channel for rare targets	Route low-prevalence rules to passive displays or batch review
Engagement is learned from experienced PPV	One unreliable rule degrades the others	Treat the portfolio, not the rule, as the unit of approval
Actioned true alerts rise monotonically with rule count	That metric cannot identify a stopping point	Do not evaluate the portfolio on caught cases alone

An optimum exists once rules differ in value	There is a defensible portfolio size	Set an explicit alert budget per clinician per day and enforce it
Higher specificity permits more rules	Quality buys capacity	Make improving an existing rule a route to deploying a new one
Base case: 3 missed and 3 false for every 1 actioned	The visible output understates the waste	Report missed and false-actioned alerts alongside catches

The fifth row is the concrete proposal. An alert budget — a stated maximum number of interruptive alerts per clinician per day, agreed in advance — converts rule approval from a sequence of independent yes-or-no decisions into a portfolio choice in which a new rule must displace an existing one or justify raising the budget. That is a governance mechanism rather than a technical one, and it is the only mechanism identified here that prevents accretion, because every individual approval decision will otherwise continue to be defensible on its own terms.

4.3 Limitations

The engagement function is the load-bearing assumption and it is not estimated from data. Its form — a hyperbolic decay in load multiplied by a power of experienced positive predictive value — is a plausible representation of two documented mechanisms, and its parameters were set so that the base configuration reproduces override rates in the reported range. Different functional forms would change the magnitudes, and possibly the location of the optima in Table 6. They would not change the direction of the specificity result, which follows from the base-rate arithmetic rather than from the behavioural model.

Rules are modelled as independent, identical and non-interacting. Real rules overlap, fire on the same patients, and vary enormously in sensitivity, specificity and target prevalence; heterogeneity of that kind would widen the range of outcomes and would generally strengthen the case for portfolio-level management rather than weaken it. Alerts are treated as uniform in disruptiveness, whereas an interruptive modal dialogue and a passive banner impose very different costs, and the distinction is exactly the one Section 4.2 recommends exploiting.

Clinicians are modelled as identical and as learning only from their own experience, with no communication between them and no institutional memory; in practice reputations of particular alerts spread socially and faster than individual experience would allow. The model also has no notion of alert content, timing or actionability, all of which are known to affect override behaviour and any of which could dominate the effects modelled here.

Finally, the value weighting in Section 3.5 is a stipulation. A harmonic decline in value was chosen for its simplicity; a steeper decline moves the optima left and a flatter one moves them right, and no empirical basis for the true shape exists. The finding to carry forward is the conditional structure — that an optimum requires declining marginal value — rather than the specific numbers 5, 5, 8 and 12.

5. CONCLUSION

A simulated portfolio of ten decision support rules at 95 % specificity produced 13.4 alerts per clinician per day with a positive predictive value of 27 %, an engagement rate of 24 %, and three missed true alerts and three false actioned alerts for every true alert acted upon. Raising specificity to 99.5 % more than doubled the useful output while cutting the load by two thirds; adding rules, by contrast, increased the catch while multiplying the waste.

Alert fatigue is therefore a design parameter rather than a clinician characteristic, and the parameter that controls it is specificity. The governance implication is that rules should be approved against a portfolio budget rather than individually, that a new rule should have to displace an existing one, and that improving the specificity of what is already deployed should be recognised as what it is: the cheapest available way to make the whole system work better, and the route by which an organisation earns the capacity to add anything new.

Declarations

Ethical approval: Not applicable.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interes: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Bashar Sabah Sahib	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

6. REFERENCES

1. H. Van Der Sijs, J. Aarts, A. Vulto, and M. Berg, "Overriding of drug safety alerts in computerized physician order entry," J Am Med Inform Assoc, vol. 13, no. 2, pp. 138-147, 2006. doi.org/10.1197/jamia.M1809
2. J. S. Ancker, A. Edwards, and S. Nosal, "Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system," BMC Med Inform Decis Mak, vol. 17, 2017. doi.org/10.1186/s12911-017-0430-8
3. K. C. Nanji, D. L. Seger, and S. P. Slight, "Medication-related clinical decision support alert overrides in inpatients," J Am Med Inform Assoc, vol. 25, no. 5, pp. 476-481, 2018. doi.org/10.1093/jamia/ocx115
4. A. D. Bryant, G. S. Fletcher, and T. H. Payne, "Drug interaction alert override rates in the Meaningful Use era: no evidence of progress," Appl Clin Inform, vol. 5, no. 3, pp. 802-813, 2014. doi.org/10.4338/ACI-2013-12-RA-0103
5. K.-G. Sl, O. Connor, and M. F. Rothschild, "Multidisciplinary approach to reducing alarm fatigue at the bedside," Crit Care Med, vol. 45, no. 9, pp. 1481-1488, 2017. doi.org/10.1097/CCM.0000000000002580
6. R. Backman, S. Bayliss, D. Moore, and I. Litchfield, "Clinical reminder alert fatigue in healthcare: a systematic literature review," Syst Rev, vol. 6, 2017. doi.org/10.1186/s13643-017-0627-z
7. P. J. Embi and A. C. Leonard, "Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study," J Am Med Inform Assoc, vol. 19, no. e1, pp. e145-e148, 2012. doi.org/10.1136/amiajnl-2011-000743
8. J. A. Osheroff, J. M. Teich, B. Middleton, E. B. Steen, A. Wright, and D. E. Detmer, "A roadmap for national action on clinical decision support," J Am Med Inform Assoc, vol. 14, no. 2, pp. 141-145, 2007. doi.org/10.1197/jamia.M2334
9. A. Wright and A. A. Ash, "Clinical decision support alert malfunctions: analysis and empirically derived taxonomy," J Am Med Inform Assoc, vol. 25, no. 5, pp. 496-506, 2018. doi.org/10.1093/jamia/ocx106
10. S. Phansalkar, H. Van Der Sijs, and A. D. Tucker, "Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records," J Am Med Inform Assoc, vol. 20, no. 3, pp. 489-493, 2013. doi.org/10.1136/amiajnl-2012-001089
11. K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," BMJ, vol. 330, no. 7494, 2005. doi.org/10.1136/bmj.38398.500764.8F

12. D. W. Bates, G. J. Kuperman, and S. Wang, "Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality," J Am Med Inform Assoc, vol. 10, no. 6, pp. 523-530, 2003. doi.org/10.1197/jamia.M1370
13. D. G. Altman and J. M. Bland, "Statistics notes: diagnostic tests 2 - predictive values," BMJ, vol. 309, no. 6947, 1994. doi.org/10.1136/bmj.309.6947.102
14. K. Goddard, A. Roudsari, and J. C. Wyatt, "Automation bias: a systematic review of frequency, effect mediators, and mitigators," J Am Med Inform Assoc, vol. 19, no. 1, pp. 121-127, 2012. doi.org/10.1136/amiajnl-2011-000089
15. Z. Co, A. J. Holmgren, and D. C. Classen, "The tradeoffs between safety and alert fatigue: data from a national evaluation of hospital medication-related clinical decision support," J Am Med Inform Assoc, vol. 27, no. 8, pp. 1252-1258, 2020. doi.org/10.1093/jamia/ocaa098
16. T. P. Morris, I. R. White, and M. J. Crowther, "Using simulation studies to evaluate statistical methods," Stat Med, vol. 38, no. 11, pp. 2074-2102, 2019. doi.org/10.1002/sim.8086