

Context-Conditioned Prescription Verification at the Point of Care: Design Principles and a Paired Evaluation of Relaya Rx

Abhishek Jha

Relaya (Vansora Studio Private Limited), India

<https://relaya.one>

Abstract. Clinical documentation software has learned to record the patient. Ambient scribes now capture pregnancy, kidney and liver function, age, weight, allergies and current medicines from the spoken encounter with high fidelity. Recording a patient’s state, however, is not the same as reasoning about it: the danger of a prescription is frequently a property not of the medicines alone but of the medicines in a particular body. We formalise prescription safety as a function of the pair (prescription, patient state), with an output space that includes an explicit *abstain* decision, and we state five measurable design criteria for a point-of-care verifier: it is right whenever it answers (C1); it never guesses, so that an invented drug, dose or finding, a silently dropped line, or a false “safe” counts as failure (C2); its certain coverage rises toward the whole prescription (C3); it is invisible to the prescriber unless it matters (C4); and its cost per prescription is low enough for population-scale use (C5). We describe Relaya Rx, a verifier built to these criteria that runs inside Relaya’s prescription workflow, and we argue that the Indian prescribing setting, with its brand-name multiplicity, fixed-dose combinations and code-mixed Hindi and English dictation, is an unusually demanding test bed for them. In a paired evaluation on 2 October 2026, nine synthetic high-risk cases were given to Relaya Rx and to a commercial AI medical scribe. Relaya Rx stopped five of the nine dangerous prescriptions before signature; the scribe stopped none and raised a warning in two. In the five cases where the danger was a property of the patient’s body, Relaya Rx identified it in four and the scribe in none. We describe a frozen, 190-case blind evaluation set and the protocol by which it will be scored.

Keywords: medication safety; clinical decision support; prescribing; pregnancy; renal impairment; hepatic impairment; selective prediction; abstention; alert fatigue; ambient clinical documentation; India; code-mixed speech

1 Introduction

Medication harm remains one of the largest preventable burdens in health care. The World Health Organization’s third Global Patient Safety Challenge, *Medication Without Harm*, set the goal of reducing severe, avoidable medication-related harm by 50% globally within five years [1]. A generation of evidence underlies that goal. *To Err Is Human* placed medical error among the leading causes of death in the United States [2]. Bates and colleagues found 6.5 adverse drug events per 100 non-obstetric admissions, 28% of them preventable, with errors arising most often at the ordering stage [3]. A meta-analysis of prospective studies estimated the incidence of serious adverse drug reactions among hospitalised patients at 6.7% and of fatal reactions at 0.32% [4]. In ambulatory care, 25% of patients receiving a prescription experienced an adverse drug event [5], and across care settings medicines account for roughly a quarter of preventable patient harm [6].

In parallel, documentation artificial intelligence has

entered clinical practice at unusual speed. Ambient AI scribes that listen to the encounter and draft the note have moved from pilot to deployment across large health systems within two years [7, 8, 9, 10], and the first randomised trial reported a reduction in documentation time with one of the two scribes tested [11]. Safety evaluations of such systems have concentrated, appropriately, on fidelity: whether the note says what was said, without hallucination or omission [12].

This paper concerns the step after fidelity. A note can faithfully record that a patient is ten weeks pregnant and, in the next section, faithfully record a prescription for an angiotensin-converting-enzyme inhibitor. Both statements are correct. The prescription is still dangerous [13, 14]. We call this *the gap between recording a patient’s state and reasoning about it*, and we argue that it is now the frontier of prescription safety at the point of care.

Relaya Rx is a prescription verifier designed for that frontier. It runs inside Relaya’s prescription workflow, between the clinician’s dictation and the signed prescrip-

tion. This report (i) formalises the problem (Section 2); (ii) states the design criteria to which Relaya Rx is built, as measurable properties rather than mechanisms (Section 3); (iii) describes the evaluation methodology, including a frozen blind case set (Section 4); (iv) reports a paired evaluation against a commercial AI medical scribe (Section 5); and (v) discusses implications and ongoing work (Sections 6 to 8). Implementation details are outside the scope of this report.

2 Problem Formulation

Prescriptions. A prescription is a finite sequence of lines

$$x = (\ell_1, \dots, \ell_n), \quad \ell_i = (d_i, q_i, f_i, \tau_i, r_i, \sigma_i), \quad (1)$$

where d_i is the medicine (resolved to its molecule or molecules, since a single brand may be a fixed-dose combination), q_i the dose with its unit, f_i the frequency, τ_i the duration, r_i the route and form, and $\sigma_i \in \{\text{NEW}, \text{CONTINUED}, \text{STOPPED}\}$ the line’s status. In practice x is not given: the clinician produces an utterance or text u , and $\hat{x}$ denotes the prescription as the verifier reads it from u .

Patient state. The patient is described by a state vector

$$s = (g, e, c, a, w, A, M, K), \quad (2)$$

comprising gestational status g (including gestational age), renal function e (for example eGFR), hepatic function c (for example Child-Pugh class), age a , weight w , the allergy set A , the set of current medicines M , and the set of active conditions K .

Danger. Let $\mathcal{H}(x, s)$ denote the set of clinically significant dangers present when prescription x is given to a patient in state s . It decomposes as

$$\mathcal{H}(x, s) = \mathcal{H}_1(x) \cup \mathcal{H}_2(x, s) \cup \mathcal{H}_s(x, s) \cup \mathcal{H}_u(x, s, h), \quad (3)$$

where $\mathcal{H}_1$ collects dangers intrinsic to a single line (a dose above an absolute ceiling, a unit error such as milligrams for micrograms), $\mathcal{H}_2$ collects pairwise dangers between two lines, or between a line and an entry of M or A (drug interactions, therapeutic duplication, drug allergy), $\mathcal{H}_s$ collects dangers that arise from the state of the patient’s body, the components g, e, c, a, w and K of s , and $\mathcal{H}_u$ collects *upstream* dangers that depend on the medication history h : prescribing cascades, in which a new medicine treats the adverse effect of an existing one [15], and withdrawal or rebound risk when a medicine is stopped.

Definition 1 (Patient-conditioned danger). *A danger δ is patient-conditioned for prescription x if there exist states s, s' such that $\delta \in \mathcal{H}(x, s)$ and $\delta \notin \mathcal{H}(x, s')$.*

The definition makes precise a familiar clinical fact: the same text is safe for one patient and harmful for another. Metformin 1000 mg twice daily is routine at

an eGFR of 80 and contraindicated at 22 [16]; glibenclamide is common in middle age and listed as potentially inappropriate in older adults because of prolonged hypoglycaemia [17]. No inspection of x alone can detect such dangers, and a system that reads x perfectly and checks every pair perfectly will still miss them unless it reasons over the state of the body.

Verification. A point-of-care verifier is a map

$$V : (u, s) \mapsto (\hat{x}, y), \quad y = (y_1, \dots, y_n), \quad y_i \in \mathcal{Y}, \quad (4)$$

with output space

$$\mathcal{Y} = \{\text{CLEAR}, \text{ABSTAIN}\} \cup \{\text{WARN}(\delta), \text{STOP}(\delta) : \delta \in \mathcal{H}(x, s)\}. \quad (5)$$

CLEAR asserts that the line is correctly read and fully checked with no significant danger; WARN(δ) and STOP(δ) report a danger δ with a recommendation to review or a requirement to resolve before signature; and ABSTAIN states, in one line to the prescriber, that the system is *not certain* about that line. The inclusion of ABSTAIN is a deliberate design choice that follows the reject-option tradition in decision theory [18] and the selective-prediction literature [19, 20]: an honest “not certain” is a valid and safe answer; a confident wrong answer is not.

3 Design Principles

We state the design of Relaya Rx as five properties, each paired with a measurement. Together they specify *what* the verifier must achieve and how that is checked, independently of how it is achieved. They derive from a single objective set by the company: an invisible system that is right in every scenario it answers and, where it cannot be certain, cannot guess.

3.1 C1: right whenever it answers

Every statement the verifier makes (medicine identity, strength, dose, frequency, duration, route, status, finding and verdict) must be correct against the case’s reference answer. Writing W for the number of wrong statements over the lines on which the verifier did not abstain,

$$W(V) = \sum_{\text{cases}} \sum_{i: y_i \neq \text{ABSTAIN}} \text{wrong}(\hat{\ell}_i, y_i), \quad (6)$$

C1 requires $W(V) = 0$. In the language of selective prediction [19, 20], C1 is a requirement of zero selective risk.

3.2 C2: never guesses

Whatever the verifier cannot establish with certainty is marked not certain, in one prescriber-facing line. Four events are outright failures: an invented medicine, an invented dose, an invented finding, and a silently dropped line. A fifth, the *false safe*, is singled out because it is

Table 1. Design criteria for a point-of-care prescription verifier, with their measurements and targets.

	Property	Measurement	Counted as failure	Target
C1	Right whenever it answers	Wrong statements per output; share of outputs with zero wrong statements	Any incorrect identity, strength, dose, frequency, duration, route, status, finding or verdict	0
C2	Never guesses	Count of each violation kind	An invented drug (wrong molecule, or a non-drug read as a drug); an invented dose or finding; a silently dropped line; a false CLEAR (“safe” for a line that is not fully checked or not certain)	0
C3	Coverage	Lines resolved and fully checked $\div$ all lines; cases with no ABSTAIN line $\div$ all cases	Never improved by trading a C2 violation for an answer	$\rightarrow 1$
C4	Invisible	Interruptions (WARN or STOP) per prescription; false interruptions on clean cases and clean lines, as count and rate	An interruption where $\mathcal{H}(x, s) = \emptyset$	false rate $\rightarrow 0$
C5	Low cost	Cost per prescription	Cost that prevents use in every consultation	low

the most dangerous output a verifier can produce:

$$FS_i = \mathbf{1} \left[y_i = \text{CLEAR} \wedge (\exists \delta \in \mathcal{H}(x, s) \text{ involving } \ell_i \vee \neg \text{checked}(\ell_i)) \right]. \quad (7)$$

C2 requires the count $G(V)$ of all five event kinds to be zero. A false safe is worse than an abstention: the first disarms the prescriber, the second alerts them.

3.3 C3: coverage

Coverage is the share of the prescription answered with certainty,

$$\varphi(V) = \frac{|\{i : y_i \neq \text{ABSTAIN}\}|}{n}, \quad (8)$$

reported per line and per case. C3 pushes φ toward 1, but always subject to C1 and C2. The criteria are therefore lexicographically ordered:

$$V^* = \arg \min_V \text{lex} (G(V), W(V), -\varphi(V), \iota_0(V), \kappa(V)), \quad (9)$$

where ι_0 and κ are defined below.

Remark. The criteria bind jointly. A verifier that always abstains satisfies C1 and C2 trivially but has $\varphi = 0$. A verifier that warns on every line catches every danger but has a false interruption rate of 1, which the alert-fatigue literature shows to be unsafe in practice [21, 22, 23]. Only the joint satisfaction of (9) describes a useful system.

3.4 C4: invisibility

Drug safety alerts in computerised order entry are over-ridden in 49% to 96% of cases [21]; in a large outpatient study, 52.6% of medication-related alerts were overridden [22], and repeated alerts for the same patient measurably reduce acceptance [23]. An interruption therefore has a cost, and an unnecessary one erodes the value

of every necessary one. We measure the interruption rate $\iota = \Pr[\exists i : y_i \in \{\text{WARN}, \text{STOP}\}]$ and, more importantly, the false interruption rate on clean prescriptions,

$$\iota_0 = \Pr[\exists i : y_i \in \{\text{WARN}, \text{STOP}\} \mid \mathcal{H}(x, s) = \emptyset]. \quad (10)$$

The target is $\iota_0 \rightarrow 0$, with no added step for the prescriber when the verifier is certain the prescription is safe. By design, the prescriber notices Relya Rx only on the prescriptions where it matters.

3.5 C5: low cost per prescription

A verifier intended for every consultation in a country as populous as India must cost very little per prescription. Cost per prescription κ is a criterion at every evaluation gate and a design constraint, not an afterthought. Relya Rx is built to a low cost per prescription.

3.6 The Indian prescribing setting as a test bed

We develop and evaluate Relya Rx in Indian outpatient practice, which we regard as an unusually demanding test bed for the criteria above, for four reasons.

Brand multiplicity. Indian prescriptions are written predominantly in brand names, and the same molecule is sold under many brands. A verifier must resolve, for example, *Levipil* to levetiracetam (not to the similarly spelled levofloxacin), *Daonil* to glibenclamide and *Thyronorm* to levothyroxine, in micrograms. A wrong resolution is a C2 failure (an invented drug), and every judgement about that line inherits the error.

Fixed-dose combinations. Fixed-dose combinations are prescribed widely in India [24]; of 118 systemic antibiotic combination formulations sold in India in one analysis, 64% had no record of central regulatory approval [25]. A single brand may contain two or three molecules, each with its own constraints in a given patient. The penicillin in *Augmentin* is the canonical example.

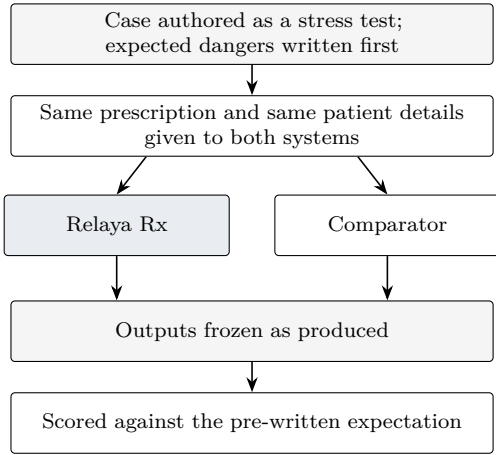


Figure 1. The paired evaluation protocol. Expectations are fixed before either system runs, and both systems see identical clinical content.

Code-mixed dictation and script. Clinicians dictate in English, Hindi, and fluid mixtures of the two, a setting in which language technology is known to be difficult [26]. A representative Hinglish instruction from our case material reads:

“Patient ko Tab Tramadol 50 mg TDS dena hai pain ke liye, aur Tab Ondansetron 4 mg SOS vomiting ke liye. Sertraline continue.”

Dictation and notes may also appear in Devanagari script, for example मरीज़ को बुखार है, “the patient has fever” (illustrative).

Error burden. A study across 13 Indian tertiary centres coordinated by the Indian Council of Medical Research found unacceptable deviations from treatment guidelines in 9.8% of 4,838 outpatient prescriptions [27], and a South Indian study found at least one medication error in 65.6% of 500 screened prescriptions [28].

We expect criteria met here to generalise. A verifier that resolves brands, combinations and code-mixed dictation correctly, and that abstains rather than guesses when it cannot, faces a simpler reading task where prescriptions use generic names and dictation is in one language, while the body-state reasoning task, $\mathcal{H}_s$, is the same in every country.

4 Evaluation Methodology

4.1 Paired protocol

Our comparative evaluations follow a paired protocol (Figure 1). (1) Each case is authored as a pharmacology stress test, and its expected dangers are written down *before either system is run*. (2) Both systems receive the same clinical content: the same prescription and the same patient details, each through the system’s normal input route. (3) Outputs are frozen as produced. (4) Each output is scored against the pre-written expectation. Because both systems see identical cases, the comparison is within-case, and differences cannot be attributed to case selection.

Table 2. Composition of the frozen blind evaluation set.

Family	Cases	Expected outcome
Extraction	70	reading only
Safety	60	42 STOP, 18 WARN
Upstream	20	cascade, withdrawal
Clean controls	40	no interruption
Total	190	
of which Hinglish	19	
of which Devanagari	5	

4.2 Endpoints

For the comparative study we report three endpoints. **E1**, dangerous prescriptions *stopped*: the system holds the prescription before signature for the prescriber’s attention. **E2**, body-state dangers *identified*: among cases whose principal danger lies in $\mathcal{H}_s$, the share in which the system names it. **E3**, specified dangers identified across all cases, scored per case with credit 1 for full identification, $\frac{1}{2}$ for partial identification and 0 otherwise. We also report the share of cases in which a system raised any safety warning. The paired study measures detection; the criterion-level measures C1 to C5 are taken on the blind set described next.

4.3 The frozen blind set

For the criterion-level evaluation of C1 to C5 we constructed a blind set of 190 cases (Table 2). It contains 70 extraction cases, in which the task is reading alone and no danger is expected; 60 safety cases, of which 42 require a stop and 18 a warning; 20 upstream cases (10 prescribing-cascade and 10 withdrawal-on-stopping cases); and 40 clean controls, prescriptions with no significant danger on which any interruption counts against C4. Nineteen cases are dictated in Hinglish and five in Devanagari. Every case carries the same schema: input text, patient state, reference record, expected findings, statements the output must not contain, and citations.

The set was frozen on 4 October 2026 by recording the SHA-256 digest of each of its files, so that any later change is detectable; the digest of that record is published in the Declarations. Builders never see it: no one who develops Relaya Rx opens it, and before scoring, all development material is searched for its case texts as a contamination check. Scoring uses the reference answer for each case; no comparison system defines the pass criterion.

5 Results: Paired Evaluation

5.1 Cases and comparator

On 2 October 2026 we ran a paired evaluation of Relaya Rx, in the configuration running inside Relaya’s prescription workflow on that date, against a commercial AI medical scribe. Ten synthetic high-risk cases were authored; one (a hyperkalaemia case) was not run on the comparator and is excluded, leaving nine paired

cases (Table 3). The cases span the danger taxonomy of Figure 2: five place the principal danger in the patient’s body (pregnancy, kidney, liver, a child’s weight, older age with kidney disease), and four are pairwise, a medicine with another medicine or with a recorded allergy. For the comparator, patient details were dictated as part of the encounter, its native input mode; for Relaya Rx they were provided as a structured patient record. The clinical content was identical. The comparison is between two products in their normal workflow, a documentation system and a verifier. Holding a prescription before signature (E1) belongs to a verifier’s workflow; naming a danger (E2, E3) is open to both, and the comparator did raise safety warnings, in two cases.

5.2 Primary results

Table 4 and Figure 3 summarise the endpoints. Relaya Rx stopped five of the nine dangerous prescriptions before signature (E1); the comparator stopped none. In the five body-state cases, Relaya Rx identified the danger in four, and the comparator in none (E2). Across all nine cases, the per-case score for specified dangers identified was 5.5 for Relaya Rx and 1.5 for the comparator (E3). The comparator raised a safety warning of any kind in two of the nine cases.

5.3 Where the difference lies

The comparator’s two warnings both fell in the pairwise class: an anticoagulant interaction in T4, which it explained clearly, and the penicillin allergy in T10, recorded as an inline caution beside a prescription that otherwise proceeded. Pairwise interaction checking is a mature category and was the comparator’s strongest; even there, Relaya Rx flagged the serotonin-syndrome risk of tramadol with sertraline in the Hinglish case (T3), which the comparator did not, and stopped the Augmentin prescription in the patient with penicillin anaphylaxis (T10), naming the allergy as a contraindication. The difference between systems concentrated where the danger is a property of the patient’s body. In each of the five body-state cases the comparator recorded the relevant state correctly in its own note (an eGFR of 22, Child-Pugh class C with prior encephalopathy, ten weeks of gestation, a weight of 14 kg, an eGFR of 25 at age 79) and then issued the prescription without comment. Relaya Rx stopped the metformin prescription at an eGFR of 22, stopped the benzodiazepine in a patient with prior hepatic encephalopathy and flagged the 4 g paracetamol dose, flagged glibenclamide at age 79 as a cause of prolonged hypoglycaemia, and stopped the pregnancy prescription described below. The case that neither system resolved was weight-based paediatric dosing (T5).

5.4 Vignette: pregnancy at ten weeks

Case T7 illustrates the gap between recording and reasoning. A 29-year-old woman, 68 kg and ten weeks pregnant, with chronic hypertension, dyslipidaemia and acne, is prescribed enalapril 5 mg once daily, atorvas-

tatin 20 mg at night, and doxycycline 100 mg twice daily for seven days. None of the three should be continued. Enalapril carries a boxed warning for fetal toxicity, and its labelling directs that it be discontinued as soon as possible once pregnancy is detected [14]; first-trimester exposure to angiotensin-converting-enzyme inhibitors has also been associated with major malformations [13]. Statins remain contraindicated in pregnancy in United Kingdom labelling [29]; United States labelling no longer lists the contraindication, following a 2021 request by the Food and Drug Administration, which still advises that most pregnant patients stop them [30]. United Kingdom guidance states that oral tetracyclines are contraindicated during pregnancy [31].

Comparator output. The scribe produced a well-structured note. Its demographic section stated the gestational status as ten weeks pregnant. Its medication table listed all three medicines with their doses and frequencies (two of them mapped to specific commercial brands that were not dictated). No warning, caution or comment accompanied any of the three.

Relaya Rx output. Relaya Rx stopped the prescription before signature. It identified enalapril as contraindicated in pregnancy, naming fetal renal dysgenesis, oligohydramnios, skull ossification defects and fetal death as the risks, and proposed pregnancy-compatible alternatives for blood pressure control. It flagged atorvastatin, to be stopped until after delivery, and doxycycline, a tetracycline, as unsafe in pregnancy. All three dangers were identified.

Both systems therefore *knew* the patient was pregnant. Only one acted on it.

6 Discussion

Body-state reasoning is the frontier. Clinical decision support has historically centred on pairwise interaction checking [32], and in the paired evaluation it was the comparator’s strongest category. Body-state dangers are different in kind. They are invisible in the prescription text, they depend on variables that change over a pregnancy or with a single creatinine result, and, by Definition 1, they cannot be detected without conditioning on *s*. Ambient documentation now carries the state spoken in the encounter into the record without additional data entry [7, 11]. The opportunity, and the responsibility, is to reason over it. Our results indicate that faithful capture of *s* is not by itself sufficient: in every body-state case the comparator’s note contained the decisive fact.

Abstention is a safety property. Most safety tooling has two outputs, an alert or silence, and silence is read as “safe”. This conflates two different states: *checked and safe*, and *not checked*. Equation (7) separates them. By making ABSTAIN a first-class output and by counting a false CLEAR as a failure, the design converts the most dangerous failure mode of a verifier, false reassurance, into the least dangerous one, a request for the clinician’s attention on one line. The cost of abstention is coverage,

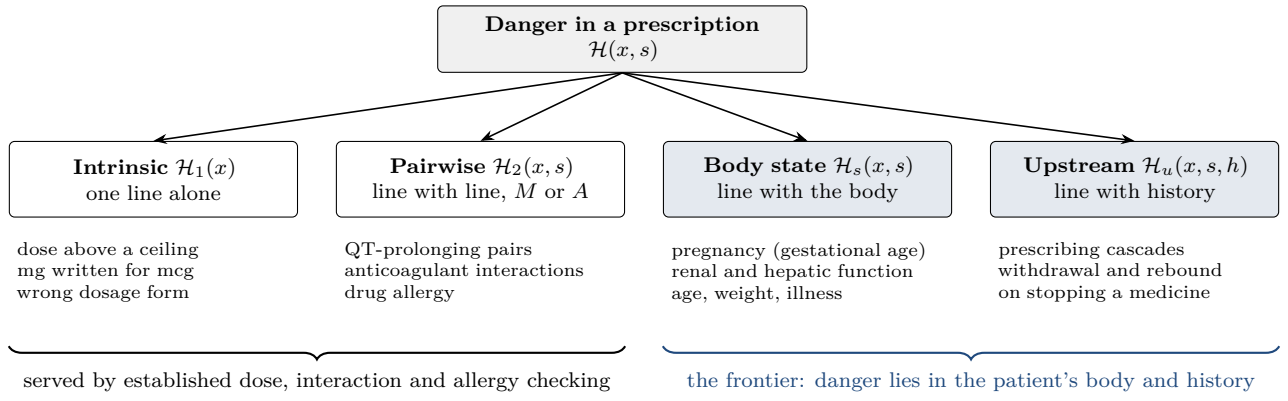


Figure 2. A taxonomy of prescription danger following Equation (3). Pairwise checking, of interactions and of drug allergy, is a mature category. Dangers conditioned on the state of the body and on medication history require reasoning over s and h ; the body-state category is where the paired evaluation found the largest difference between systems.

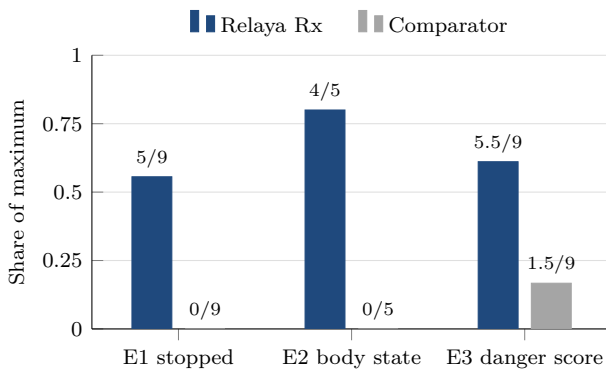


Figure 3. Endpoints of the paired evaluation. Labels give the counts; bar height is the share of the maximum, out of 9 cases for E1 and E3 and out of 5 for E2.

which is measured (C3) and driven upward; the cost of a guess is a patient.

Invisibility is a safety property too. The override literature [21, 22, 23] shows that the marginal alert has negative value once alerts become frequent. A verifier that is right whenever it answers, and silent when the prescription is clean, preserves the signal value of every interruption it does make. C4 is accordingly measured on clean controls, not only on dangerous cases.

What population-scale use requires. Universal verification, in which every prescription in every consultation is checked against the patient, is a different engineering target from selective decision support for hospital order entry. It requires reading in the language and script the clinician actually uses, resolving the brands and combinations actually prescribed, adding no step for the prescriber when the prescription is safe, and costing little enough per prescription to be offered everywhere. These are the requirements behind C1 to C5, and they are why we treat the Indian outpatient setting as the proving ground.

7 Ongoing Work

Blind-set evaluation. The 190-case blind set of Section 4 is frozen and is the next scheduled measurement. It will report C1 to C4 with denominators, including the false interruption rate on the 40 clean controls; C5, cost per prescription, is a constraint at every gate.

Coverage toward the whole prescription. C3 is being driven toward 1 across brand resolution, combination products, code-mixed dictation and Devanagari input, and no evaluation gate is passed unless C1 and C2 are zero on every case.

Upstream reasoning. The upstream family $\mathcal{H}_u$, comprising prescribing cascades [15] and withdrawal or rebound risk when a medicine is stopped, extends verification from the current prescription to the patient's medication history. Twenty of the blind cases test it directly.

Weight-normalised paediatric dosing. The one paired case that neither system resolved, a child's dose expressed per kilogram, is an active target and is represented in the blind set.

8 Limitations

The paired evaluation used synthetic cases written by the developer's team, which also scored the outputs and was not blinded to system identity. The sample is small ($n = 9$), each system was run once per case, and no patient outcomes were measured. The paired study scored the specified dangers; it did not score every other statement in each output, which the blind set measures under C1 and C2. The comparator was assessed on its generated note; notices displayed transiently on screen but not carried into the note were not captured. Results describe each system as configured on the test date. The blind-set evaluation of Section 4 enlarges the sample from 9 to 190 cases, adds 40 clean controls, and scores every output against a reference answer frozen with the cases.

Table 3. The nine paired cases. All are synthetic; expected dangers were written before either system ran. Case identifiers follow the original authoring sequence.

Case	Class	Patient state	Prescription and expected danger
T1	Pairwise	67 y, male; atrial fibrillation on amiodarone; pneumonia	Clarithromycin with amiodarone: QT prolongation, raised amiodarone exposure
T2	Body state	76 y, female; CKD stage 4, eGFR 22	Metformin 1000 mg twice daily contraindicated below eGFR 30; gabapentin and ciprofloxacin need renal dose reduction
T3	Pairwise	44 y, female; on sertraline	Dictated in Hinglish. Tramadol with sertraline: serotonin syndrome and seizure risk; ondansetron adds serotonergic load
T4	Pairwise	59 y, female; mechanical mitral valve on warfarin	Fluconazole and metronidazole both potentiate warfarin: INR rise, bleeding
T5	Body state	4 y, male; 14 kg; fever	Paracetamol syrup at 143 mg/kg/day and ibuprofen at 43 mg/kg/day, both above weight-based maxima
T6	Body state	58 y, male; cirrhosis, Child-Pugh C, prior hepatic encephalopathy	Alprazolam may precipitate encephalopathy; paracetamol 4 g/day exceeds the hepatic ceiling
T7	Body state	29 y, female; 10 weeks pregnant; hypertension, dyslipidaemia, acne	Enalapril (fetal toxicity), atorvastatin and doxycycline: none to be continued in pregnancy
T9	Body state	79 y, male; CKD stage 4, eGFR 25	Indian brand names. Daonil (glibenclamide): prolonged hypoglycaemia in older adults; Levipil (levetiracetam) needs renal dose reduction
T10	Pairwise	52 y, female; penicillin anaphylaxis; on methotrexate	Augmentin contains amoxicillin; amoxicillin and diclofenac reduce methotrexate clearance

Table 4. Paired evaluation, 2 October 2026 ($n = 9$). The comparator raised a safety warning of any kind in 2 of 9 cases.

Endpoint	Relaya Rx	Comparator
E1 Prescriptions stopped	5/9	0/9
E2 Body-state dangers found	4/5	0/5
E3 Dangers identified (score)	5.5/9	1.5/9

Declarations

Competing interests. The author is the founder of Relaya, the developer of Relaya Rx. **Funding.** Vansora Studio Private Limited. **Ethics.** All cases are synthetic; no patient data were used. **Data availability.** The paired cases are described in Table 3. Blind-set contents are withheld to preserve blindness. The SHA-256 digests of its files were recorded at freezing on 4 October 2026; the SHA-256 digest of that record is 66195ff9 c41b58ab c8ce9908 5ecf6e2c e6d2a3e7 c2ef1ceb 3bc6f670 0a565997.

References

- [1] World Health Organization. *Medication Without Harm: WHO Global Patient Safety Challenge on Medication Safety*. Geneva: World Health Organization; 2017. WHO/HIS/SDS/2017.6.
- [2] Institute of Medicine; Kohn LT, Corrigan JM, Donaldson MS, editors. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000. doi:10.17226/9728.
- [3] Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events: implications for prevention. *JAMA*. 1995;274(1):29-34. doi:10.1001/jama.1995.03530010043033.
- [4] Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*. 1998;279(15):1200-1205. doi:10.1001/jama.279.15.1200.
- [5] Gandhi TK, Weingart SN, Borus J, et al. Adverse drug events in ambulatory care. *N Engl J Med*. 2003;348(16):1556-1564. doi:10.1056/NEJMsa020703.
- [6] Panagioti M, Khan K, Keers RN, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ*. 2019;366:l4185. doi:10.1136/bmj.l4185.
- [7] Tierney AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catalyst*. 2024;5(3). doi:10.1056/CAT.23.0404.
- [8] Liu TL, Hetherington TC, Dharod A, et al. Does AI-powered clinical documentation enhance clinician efficiency? A longitudinal study. *NEJM AI*. 2024;1(12). doi:10.1056/AIoa2400659.
- [9] Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. *J Am Med Inform Assoc*. 2025;32(2):375-380. doi:10.1093/jamia/ocae295.
- [10] Duggan MJ, Gervase J, Schoenbaum A, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. *JAMA Netw Open*. 2025;8(2):e2460637. doi:10.1001/jamanetworkopen.2024.60637.
- [11] Lukac PJ, Turner W, Vangala S, et al. Ambient AI scribes in clinical practice: a randomized trial. *NEJM AI*. 2025;2(12). doi:10.1056/AIoa2501000.

- [12] Asgari E, Montaña-Brown N, Dubois M, et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *npj Digit Med*. 2025;8:274. doi:10.1038/s41746-025-01670-7.
- [13] Cooper WO, Hernandez-Diaz S, Arbogast PG, et al. Major congenital malformations after first-trimester exposure to ACE inhibitors. *N Engl J Med*. 2006;354(23):2443-2451. doi:10.1056/NEJMoa055202.
- [14] Bausch Health US. *Vasotec (enalapril maleate) tablets*: prescribing information; boxed warning, fetal toxicity. DailyMed, US National Library of Medicine. Accessed 5 October 2026.
- [15] Rochon PA, Gurwitz JH. Optimising drug treatment for elderly people: the prescribing cascade. *BMJ*. 1997;315(7115):1096-1099. doi:10.1136/bmj.315.7115.1096.
- [16] Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2022 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney Int*. 2022;102(5S):S1-S127. doi:10.1016/j.kint.2022.06.008.
- [17] 2023 American Geriatrics Society Beers Criteria Update Expert Panel. American Geriatrics Society 2023 updated AGS Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2023;71(7):2052-2081. doi:10.1111/jgs.18372.
- [18] Chow CK. On optimum recognition error and reject tradeoff. *IEEE Trans Inf Theory*. 1970;16(1):41-46. doi:10.1109/TIT.1970.1054406.
- [19] El-Yaniv R, Wiener Y. On the foundations of noise-free selective classification. *J Mach Learn Res*. 2010;11:1605-1641.
- [20] Herbei R, Wegkamp MH. Classification with reject option. *Can J Stat*. 2006;34(4):709-721. doi:10.1002/cjs.5550340410.
- [21] van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc*. 2006;13(2):138-147. doi:10.1197/jamia.M1809.
- [22] Nanji KC, Slight SP, Seger DL, et al. Overrides of medication-related clinical decision support alerts in outpatients. *J Am Med Inform Assoc*. 2014;21(3):487-491. doi:10.1136/amiajnl-2013-001813.
- [23] Ancker JS, Edwards A, Nosal S, et al.; HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Med Inform Decis Mak*. 2017;17:36. doi:10.1186/s12911-017-0430-8.
- [24] Gupta YK, Ramachandran SS. Fixed dose drug combinations: issues and challenges in India. *Indian J Pharmacol*. 2016;48(4):347-349. doi:10.4103/0253-7613.186200.
- [25] McGettigan P, Roderick P, Kadam A, Pollock A. Threats to global antimicrobial resistance control: centrally approved and unapproved antibiotic formulations sold in India. *Br J Clin Pharmacol*. 2019;85(1):59-70. doi:10.1111/bcp.13503.
- [26] Khanuja S, Dandapat S, Srinivasan A, Sitaram S, Choudhury M. GLUECoS: an evaluation benchmark for code-switched NLP. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020:3575-3585. doi:10.18653/v1/2020.acl-main.329.
- [27] Shetty Y, Kamat S, Tripathi R, et al. Evaluation of prescriptions from tertiary care hospitals across India for deviations from treatment guidelines and their potential consequences. *Indian J Med Res*. 2024;159(2):130-141. doi:10.4103/ijmr.ijmr_2309_22.
- [28] Kandasamy G, Sivanandy P, Almaghaslah D, et al. A cross-sectional study on prescribing and dispensing errors at a corporate hospital in South India. *Int J Clin Pract*. 2021;75(9):e14489. doi:10.1111/ijcp.14489.
- [29] Viatris Products Limited. *Lipitor 20 mg film-coated tablets*: summary of product characteristics, section 4.3, contraindications. Electronic Medicines Compendium; text revised November 2025. Accessed 5 October 2026.
- [30] US Food and Drug Administration. FDA requests removal of strongest warning against using cholesterol-lowering statins during pregnancy; still advises most pregnant patients should stop taking statins. Drug Safety Communication; 20 July 2021.
- [31] National Institute for Health and Care Excellence. *Acne vulgaris: management*. NICE guideline NG198. London: NICE; 2021. Recommendation 1.5.8.
- [32] Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digit Med*. 2020;3:17. doi:10.1038/s41746-020-0221-y.