

Live ,the Liver Life: From equations to allocation: a wave-form prior for early planning of hepatitis C screening under NVHCP and the WHO 2030 strategy

Shrish Chandra Srivastava

shrish.093@gmail.com

Banaras Hindu University <https://orcid.org/0009-0003-2971-762X>

Research Article

Keywords: hepatitis C, NVHCP, WHO elimination, mathematical epidemiology, Bayesian prior, screening allocation, Eastern Uttar Pradesh, Varanasi, machine learning

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11348675/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

From equations to allocation: a wave-form prior for early planning of hepatitis C screening under NVHCP and the WHO 2030 strategy

Twelve equations from The Synthesis of Infectious Diseases, Part Two; a pre-declared prediction for the Model Treatment Centre, BHU, Varanasi; and the conditions under which machine learning would improve it

Shrish Chandra

Model Treatment Centre, Banaras Hindu University, Varanasi, Eastern Uttar Pradesh, India

Status. Preprint, not peer reviewed. The central result is a *prediction* and the supporting analyses are *simulations*. No primary data were collected, no patient was screened, and nothing here is clinical guidance.

Abstract

Background. India's National Viral Hepatitis Control Programme (NVHCP) aims at countrywide elimination of hepatitis C by 2030, and the WHO strategy sets service targets of 90% of infections diagnosed and 80% of eligible patients treated. Planning reagents, direct-acting antivirals (DAAs) and staff for a district requires a prevalence, which is slow and costly to measure and which meets participation resistance. Eastern Uttar Pradesh has no published primary series.

Methods. Six of the twelve equations of Part Two were computed (Eq 1, 2, 4, 8, 10, 11) and two were used structurally (Eq 3, 5); four were not used. The corrected latitudinal envelope $\Phi_{\text{env}}(\lambda) = \cos \lambda$ was used to transport the corpus's anchor prevalence (11.6%, Lucknow) to Varanasi. The resulting band was converted into a Beta prior, and three simulations quantified (i) the sample-size value and the coverage risk of that prior, (ii) the yield of targeted over uniform screening as a function of predictor skill, and (iii) the share of prediction uncertainty carried by latitude.

Results. Predicted HCV RNA-positive prevalence among gastroenterology OPD attendees at the Model Treatment Centre, BHU, is **11.33%** (band 9.06–13.60%), equivalent to one case per 8.8 attendees screened (7.4–11.0). The framework prior is worth about **782** screened attendees: it reduces the sample needed for a ± 1 percentage-point interval from 3,861 to 3,076 (20%) and for a ± 2 point interval from 967 to 181 (81%). Used at nominal weight it is unsafe: at a true prevalence of 7% the 95% interval covers the truth in only 6% of replicates. Weakening the prior fourfold keeps coverage at 87% or above from 7% to 16% true prevalence. Across all of Eastern Uttar Pradesh the latitudinal envelope moves the prediction by 0.16 percentage points, against 4.53 for the unmeasured incidence ratio. Targeting screening by a predictor of district prevalence raises cases found per screen to 1.33 times the uniform-allocation yield ($\sigma = 0.5$, $\rho = 0.5$, top-quartile rule), but the latitude envelope alone could contribute at most 0.3%.

Conclusions. The framework is useful as a *starting prior and a specification*, not as a district-level map. It lets planning begin before data exist, quantifies what that head start is worth, and states the predictor skill a machine-learning layer must reach before it should alter any allocation. Four pre-declared tests (F6–F9) gate each step.

Keywords: hepatitis C; NVHCP; WHO elimination; mathematical epidemiology; Bayesian prior; screening allocation; Eastern Uttar Pradesh; Varanasi; machine learning

1. Introduction

India launched the National Viral Hepatitis Control Programme on 28 July 2018, with the aim of countrywide elimination of hepatitis C by 2030 [1]. The Government of India estimated 0.6–1.2 crore (6–12 million) people living with hepatitis C at launch [2]. The WHO strategy defines elimination as a 90% reduction in incidence and a 65% reduction in mortality [3,4], with 2030 service targets of 90% of infections diagnosed and 80% of eligible patients treated [3,4]. Globally, by the end of 2022 only 36.4% of people living with HCV had been diagnosed and 20% treated [6]. Within India the weakest step is often after detection: in one NVHCP apex-centre study of blood donors, only 8.9% of anti-HCV-positive donors attended the NVHCP clinic [5].

A programme that must stock reagents and DAAs, and staff a clinic, before a district has been surveyed needs an expected prevalence. Measuring it is the obstacle. Population surveys are slow, consenting participants and returning them for confirmatory RNA testing meets resistance, and the target is a moving cascade. This work grew from exactly that obstacle: a primary prevalence study at the gastroenterology OPD of the BHU Model Treatment Centre, designed to screen attendees, record risk factors, and follow viraemic patients by PCR to sustained virological response, was left unfinished, and a mathematical framework was built in its place.

The framework is the twelve-equation wave-form theory of *The Synthesis of Infectious Diseases, Part Two* [7]. The computational audit of its first edition [8], and the derivation in the companion finding [9], indicate that the framework consumes an anchor prevalence rather than producing one. This paper therefore asks a narrower and answerable question than "can the equations replace a survey": **what planning value can an equation-derived prior carry, how much is it worth in screened people, where does it fail, and what would a machine-learning layer need to achieve before it changed an allocation?**

The contributions are: (i) an explicit enumeration of which of the twelve equations were computed, which were used structurally, and which were not used (Table 1); (ii) a pre-declared prediction for one named clinic (§4.1); (iii) a quantified value-and-risk analysis of the framework prior (§4.3); (iv) a simulation of the predictor skill required for targeted screening to pay (§4.4); and (v) a validation register that gates each extension (§8).

Figure 1 gives the staged logic used throughout.

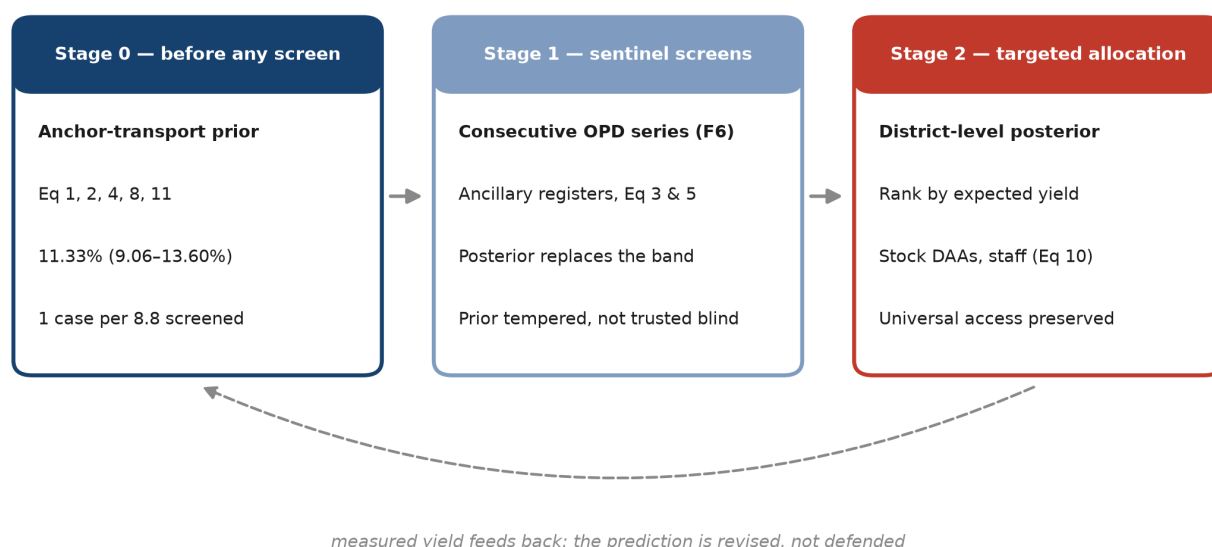


Figure 1. Staged use of the framework. Stage 0 needs no local data and rests on Equations 1, 2, 4, 8 and 11. Stage 1 replaces the prior band with a posterior from sentinel screens, using ancillary registers structured by Equations 3 and 5, with the prior tempered until test F6 passes. Stage 2 ranks districts by expected yield and sizes DAA stock and staff against Equation 10 and the WHO service targets. Measured yield feeds back into the prediction.

2. The framework and the equations used

Table 1 enumerates all twelve equations of Part Two, the role each played here, and its standing in the author's own audit registers (§15.8 and §15.9 of Part Two). An earlier draft of this work counted five equations; Equation 10 was also computed, and the count is six.

Eq	Name and form	Role in this paper	Standing in Part Two
1	P-I-D identity: $P = I \times D^-$	Computed. Yields prevalence from incidence and duration	Identity
2	Wave frequency: $f = 1/D^- = I/P$	Computed. Frequency is inverse duration	Identity
3	Numerator: $I = \sum J_{\text{route}}$ (nine flux terms)	Structural. Defines the feature space of the ML extension (§7)	Three differing rosters; no vertical-transmission term (R8, owed)
4	Denominator: $D_{\text{op}}^- = D_{\text{biol}} / \prod D_i$ (eight stock terms)	Computed. Operational duration; exposes defect IS-1c	Cannot represent curative therapy: range is $[D_{\text{biol}}, \infty)$
5	Full frequency: $f_k(\phi, d) = \sum J / (P_k \cdot D_{\text{op}}^-)$	Structural. The cell-level model into which data are assimilated	None recorded
6	Wave energy: $\Psi = \frac{1}{2} \kappa A^2 f^2$	Not used	Defined three ways (R5, owed)
7	Wave ratio: Ψ_{k+1} / Ψ_k	Not used	Postdicted (F3)
8	Latitudinal envelope: $ID = f_{\text{max}} \cdot \cos \lambda \cdot \Phi(\phi, d, t)$	Computed. Transport between cells	Corrected at R1; one table and one passage still print the withdrawn form
9	Dissipation: $\delta = \text{fatality ratio} / \text{wave ratio}$	Not used	Three incompatible definitions (R3); circular with Ψ (R5)
10	Elimination threshold: $\delta \geq 0.40$	Computed. Programme scaling, coverage definition	Near-tautology under Eq 9; sharp only under the coverage definition (R3, R4)
11	Persistence index: $\Pi = \exp(-f)$	Computed. Persistence of chronic HCV	Identity; cannot fail
12	Global wave formula: $\Psi_{\text{global}} = \sum \kappa v^2 A^2 \delta_{\text{cycle}}$	Not used computationally; atlas role in §7	Depends on R4 and R5; assumption AS-2 contradicted by J_{mob} (R10)

Two properties of this table govern everything that follows. First, only Equations 1, 2 and 11 are identities; they are reliable because they cannot fail, and for the same reason they carry no empirical content. Second, the only equation that *adds geography* is Equation 8, and its geographic term is $\Phi_{\text{env}}(\lambda) = \cos \lambda$, normalised at the pole after the repair recorded at R1 of Part Two. The first-edition normalisation at $\lambda_{\text{max}} = 23.4^\circ$ returns a negative value at Varanasi (-0.126) and is not used.

3. Methods

3.1 Transport derivation

Equation 8 modulates ID, which Part Two defines as inverse duration, that is, wave frequency. Equation 1 then gives $P = I / f$. Prevalence is therefore proportional to $I / \cos \lambda$: the envelope enters prevalence *inversely*. Transporting the corpus's Lucknow anchor (Box C, 28.0° N, 11.6%) to Varanasi (25.3176° N) at equal incidence gives

$$\begin{aligned} P(\text{Varanasi}) / P(\text{Lucknow}) &= [I(\text{Varanasi}) / I(\text{Lucknow})] \times \cos(28.0000^\circ) / \cos(25.3176^\circ) \\ &= [I(\text{Varanasi}) / I(\text{Lucknow})] \times 0.9768 \end{aligned}$$

The incidence ratio $I(\text{Varanasi}) / I(\text{Lucknow})$ is the single unmeasured parameter. It was swept from 0.80 to 1.20, giving the planning band. The anchor's denominator is not stated in the corpus; an 11–12% value is an order of magnitude above India's community figure, so the anchor is a clinic or high-risk series, and the prediction is a *tertiary-referral-frame* prediction.

3.2 From a band to a prior

The band is a sweep over an unmeasured parameter, not a confidence interval. Converting it to a probability distribution is therefore a **declared convention**: the band is read as ± 2 standard deviations around the point value, giving a standard deviation of 1.13 percentage points and a Beta(88.6, 693.0) prior with effective sample size 782. A *tempered* prior divides both Beta parameters by four. A *mixture* prior gives equal weight to the framework Beta and a flat Beta(1,1), updated by marginal likelihood.

3.3 Value and risk of the prior

For each prior, the number of consecutive attendees needed for a 95% equal-tailed interval of half-width h was found by integer bisection at the expected count ($k = n \times 11.33\%$). Coverage was estimated by simulation: 1,500 replicate screens of $n = 1,000$ at each of seven true prevalences (5, 7, 9, 11.33, 14, 16, 20%), recording how often the interval contained the truth, its mean width and its bias. Random seed 20261012.

3.4 Allocation simulation

$K = 28$ hypothetical districts were given true prevalences $p_k = \exp(\mu + \sigma z_k)$ with z_k standard normal; $\sigma \in \{0.25, 0.5, 0.75\}$ sets how much districts truly differ. A predictor was constructed with correlation ρ to the true standardised district effect. Two rules allocated a fixed number of screens: *proportional* to predicted prevalence, and *top quartile only* (equal screens to the seven highest-ranked districts). The outcome is cases found per screen relative to uniform allocation, averaged over 4,000 replicates. K and σ are parameters, not claims about Eastern Uttar Pradesh. The ceiling on the skill of the latitude envelope alone was bounded by assuming its sign were correct and all of its variance were signal: $\rho_{\max} = \text{sd}(\ln \cos \lambda) / \sigma$ across the region's 25.146° – 26.814° N span.

3.5 Planning arithmetic

Number needed to screen is $1/p$. Expected RNA-positive cases per n screened is $n \times p$. DAA initiations were scaled at two coverage fractions of the viraemic pool: 0.40 (Equation 10, coverage definition) and 0.80 (the WHO service target for eligible patients [3,4]). All quantities were computed by `paper_calc.py`.

4. Results

4.1 Prediction and planning arithmetic

Predicted HCV RNA-positive prevalence among patients presenting to the gastroenterology OPD, Model Treatment Centre, BHU, Varanasi: 11.33% (planning band 9.06–13.60%). Table 2 gives the sensitivity to the unmeasured incidence ratio.

$I(\text{Varanasi}) / I(\text{Lucknow})$	Predicted prevalence	Attendees screened per case
0.80	9.06%	11.0
0.90	10.20%	9.8
1.00	11.33%	8.8

I(Varanasi) / I(Lucknow)	Predicted prevalence	Attendees screened per case
1.10	12.46%	8.0
1.20	13.60%	7.4

Table 3 converts the prediction into planning quantities per screened cohort.

Attendees screened	Expected RNA-positive (band)	DAA initiations at coverage 0.40	DAA initiations at coverage 0.80
1,000	113 (91–136)	45	91
5,000	567 (453–680)	227	453
10,000	1,133 (906–1,360)	453	906

Two cautions apply to Table 3. First, the 0.40 column is the framework's own criterion and is *half* the WHO service target; a programme planning against WHO targets should use the 0.80 column. Second, these are counts of viraemic people found *within a screened cohort*, before linkage losses. In a treatment-centre OPD linkage is immediate; for community screening it is not [5].

Frame discipline. The clinic-frame prediction must not be applied to community screening. At the corpus's all-India community figure of 0.85% the number needed to screen is 118, so using the clinic-frame figure of 8.8 would overstate community yield by a factor of 13.

4.2 Where the uncertainty lives

Across the whole latitudinal span of Eastern Uttar Pradesh (Mirzapur 25.15° N to Basti 26.81° N) the envelope $\cos \lambda$ moves from 0.9052 to 0.8925, a 1.4% range, and shifts the prediction by 0.16 percentage points. The incidence ratio, swept over $\pm 20\%$, shifts it by 4.53 percentage points, a 28-fold larger range (Figure 2). This is a comparison of ranges, not of variances.

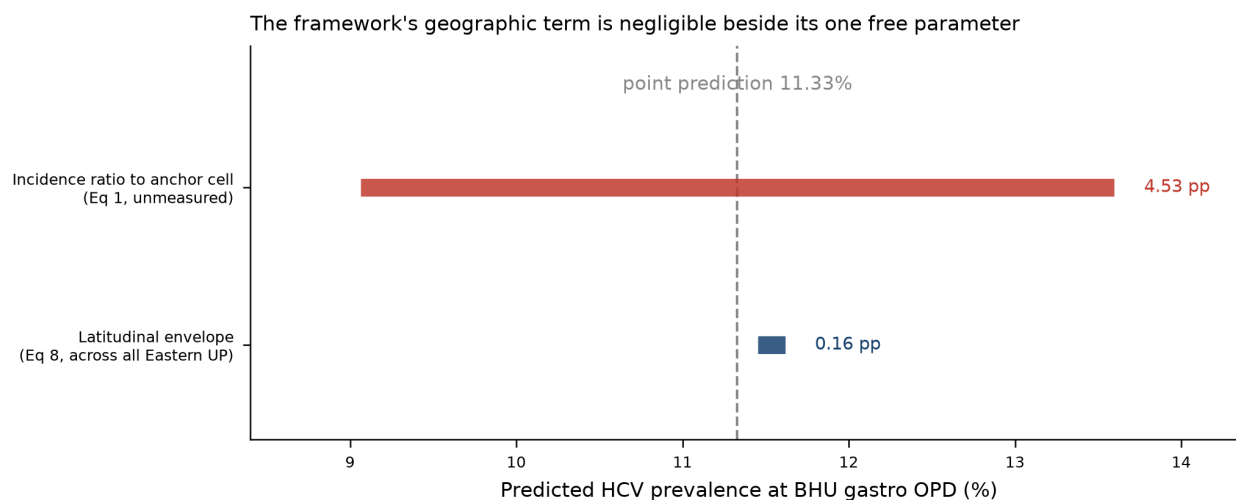


Figure 2. Share of prediction uncertainty carried by each term. Across the whole latitudinal span of Eastern Uttar Pradesh the envelope (Eq 8) moves the prediction by 0.16 percentage points; the unmeasured incidence ratio (Eq 1), swept over $\pm 20\%$, moves it by 4.53, a 28-fold larger range. The dashed line is the point prediction.

4.3 Value of the prior, and its risk

Table 4 gives the consecutive attendees needed for a 95% interval of the stated half-width.

Half-width	Flat prior	Framework, nominal	Framework, tempered	Mixture
± 2 pp	967	181	767	301
± 1 pp	3,861	3,076	3,662	3,107
± 0.5 pp	15,440	14,654	15,240	14,683

The saving at nominal weight is constant at about 782 screens, as conjugate theory requires, so its *relative* value is largest where precision demands are coarse: 81% at ± 2 pp, 20% at ± 1 pp and 5% at ± 0.5 pp. The flat-prior figure at ± 1 pp agrees with the Wald formula (3,860). Planning-grade precision, ± 2 pp, is where the framework head start matters.

The same prior is dangerous when wrong (Figure 3b). At nominal weight and $n = 1,000$, coverage of the 95% interval is 0% at a true prevalence of 5%, 6% at 7%, 77% at 9% (inside the framework's own band), 75% at 14%, and 28% at 16%. The interval is narrower than the flat-prior interval and misses the truth more often. The tempered prior stays at 87% or above from 7% to 16% true prevalence, at a cost of a smaller saving (5% at ± 1 pp). The mixture prior protects the extremes but under-covers at 9% (83%) and 14% (82%), the shoulders of the band, probably because at $n = 1,000$ the marginal likelihood still favours the informative component (this explanation was not tested separately).

Consequence. The framework prior should be used at nominal weight only after the sentinel screen (F6, §8) has shown that the observed prevalence lies inside the band; until then the tempered prior is the defensible default.

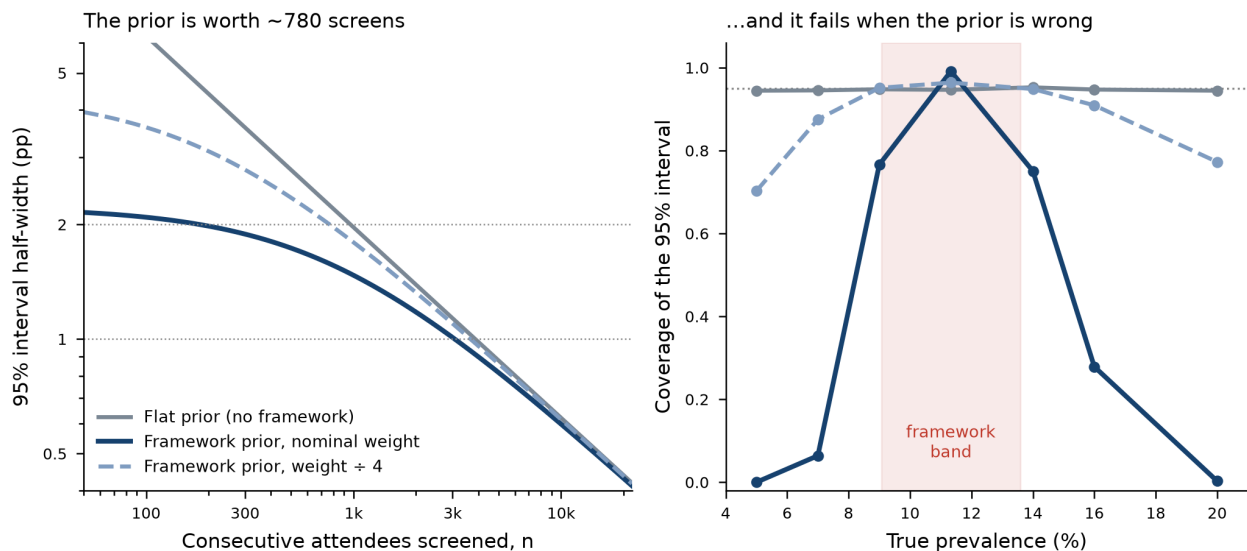


Figure 3. Value and risk of the framework prior. (a) Half-width of the 95% interval against consecutive attendees screened, at the expected count under the point prediction; the nominal prior is worth about 780 screens. (b) Coverage of the 95% interval at $n = 1,000$ (1,500 replicates per point) across true prevalences, with the framework band shaded and the 95% level dotted. Colours and line styles are shared between panels.

4.4 Targeting: how much predictor skill is enough

Figure 4 shows cases found per screen, relative to uniform allocation, as a function of predictor skill ρ . Gains depend jointly on skill and on how much districts truly differ (σ). At $\sigma = 0.5$, a predictor with $\rho = 0.5$ raises yield to 1.12 times the uniform-allocation yield under proportional allocation and 1.33 times under the top-quartile rule; at $\rho = 0.8$ the figures are 1.21 and 1.54 times. If districts differ more ($\sigma = 0.75$) the same skill ($\rho = 0.8$) gives 1.50 and 1.83 times; if they differ little ($\sigma = 0.25$), only 1.05 and 1.26 times. For the top-quartile rule at $\sigma = 0.5$ and $\rho = 0.5$, the number of people who must be approached per case found falls by 25%.

The latitude envelope alone cannot deliver any of this. Its log-spread across Eastern Uttar Pradesh has standard deviation 0.0041. Even if its sign were right and all of its variance were signal, the achievable correlation is $\rho \leq 0.016$ ($\sigma = 0.25$) and 0.008 ($\sigma = 0.5$), and the yield gain is at most 0.3%. A district-level bird's-eye view therefore cannot come from Equation 8 within a state; it has to come from the flux terms (Eq 3) and stock terms (Eq 4) populated with data.

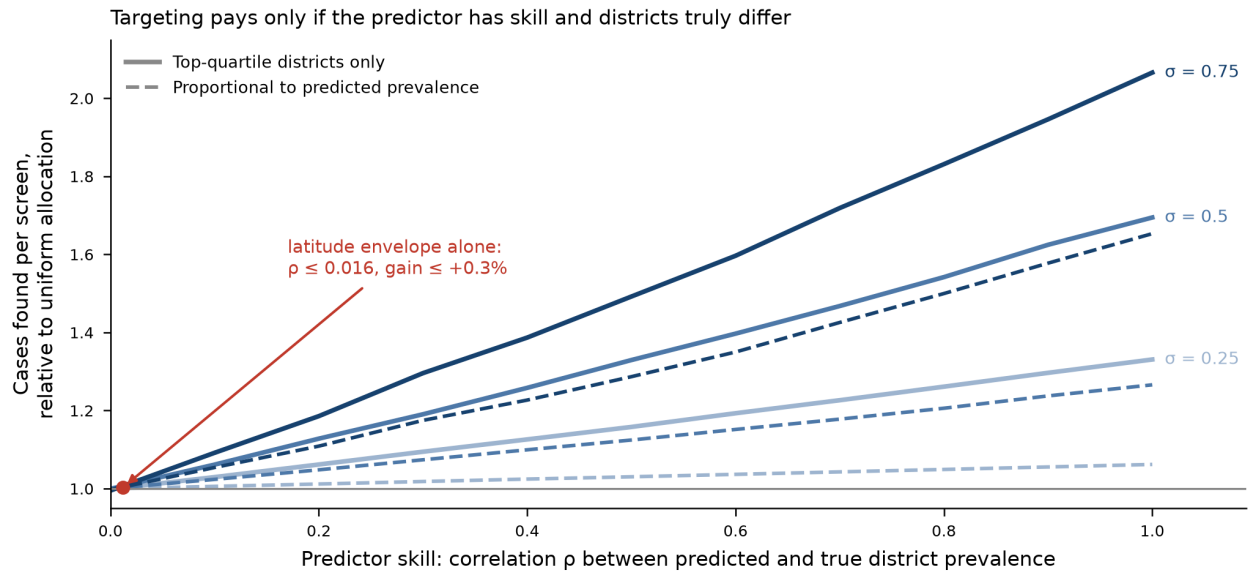


Figure 4. Cases found per screen relative to uniform allocation, against predictor skill ρ , for $K = 28$ hypothetical districts and three levels of true between-district heterogeneity σ (4,000 replicates per point). Solid: screens concentrated on the top-quartile districts. Dashed: screens proportional to predicted prevalence. The red point marks the best case for the latitude envelope alone.

4.5 Structural defects exposed by running the chain

Running the chain for one named clinic exposed three properties of the equations, reported in the companion finding [9]. **IS-1a:** the envelope enters prevalence inversely, whereas the Disease 1 panel of Part Two applies the 0.9338 frequency ratio directly to prevalence; the prevalence ratio the equations require is 1.0709. **IS-1b:** on the corpus's own two anchored cells (Mumbai 12.4%, Lucknow 11.6%) the observed Lucknow-to-Mumbai prevalence ratio is 0.9355, opposite in direction to 1.0709, so the envelope must be overcome by a 13% lower incidence in Lucknow. **IS-1c:** Equation 4 divides by a product of terms in $(0,1]$, so $D_{op}^- \geq D_{biol}$ and curative treatment, a 12-week duration, would need $\prod D_i = 40$. The operational denominator can describe a weak health system but not a working one. These must be repaired before Equations 4 and 8 are used to forecast the effect of an intervention.

5. Implications for NVHCP and WHO planning

5.1 Early access

Before any local screen, the prior supports four planning actions: sizing laboratory capacity (expected positives per 1,000 screened, Table 3), pre-positioning DAA stock against the 0.80 column, setting the sample for the first sentinel series, and fixing the F6 acceptance interval. At planning-grade precision (± 2 pp) the nominal-weight prior reduces the first sentinel sample from 967 to 181 consecutive attendees. That interval is, however, mostly the prior speaking (§4.3), so it is a planning figure and not a prevalence estimate.

5.2 Where to look

Equation 8 can discriminate between locations only when they span a wide latitude range. Mainland India spans roughly 8° N to 35° N, over which $\cos \lambda$ changes by about 17%, against 1.4% across Eastern Uttar Pradesh. Even there, IS-1b shows the sign is not established on the only two anchored points. The envelope is therefore a coarse inter-regional prior, not a district locator. The active-case cohort is found by the framework's *numerator terms*: parenteral exposure J_{par} , sexual exposure J_{sex} , mobility J_{mob} (Eastern Uttar Pradesh is a region of substantial out-migration) and sex differential J_{dif} , each of which requires a data stream.

5.3 Budget

The staged design in Figure 1 defers the large allocation decision until a measurement exists. A proportion of the budget that would otherwise be spent on blanket mobilisation can be redirected to active case-finding where the model predicts the highest yield, but §4.4 shows the gain depends on two quantities that no equation supplies: how much districts truly differ (σ) and how much skill the predictor has (ρ). Both can be estimated cheaply from a first round of sentinel screens before any budget moves. Targeting is a rule for *prioritising active case-finding*. It must not be used to ration access: WHO and NVHCP policy is to treat all diagnosed patients, and districts with low predicted prevalence keep standard access.

5.4 Participation and resistance

The method reduces the number of people who must be approached in two measurable ways: the prior is worth 782 screened attendees for a given precision (20% of the sample at ± 1 pp), and targeting reduces approaches per case found (§4.4). It does not remove the need for informed consent, does not address stigma, and does not address the linkage gap that follows detection [5]. A programme that finds more cases per screen and then, as in the cited blood-donor series [5], sees 91% of reactive donors never reach the clinic has moved its bottleneck, not removed it.

6. Extension to other geographies and diseases

Equation 3 of Part Two assigns each disease a set of dominant transmission routes (Table 5). The same chain, Equation 1 for the identity, Equation 4 for the health-system stocks, Equation 8 for transport, applies to any disease for which an anchor cell and a mean duration exist.

Route term	Dominant diseases in Part Two
J_{sex}	HIV, HBV, HCV (genotype 1b)
J_{par}	HCV (genotype 3a, India), HBV, HIV
J_{vec}	Dengue, Zika, Chikungunya, Japanese encephalitis, malaria
J_{resp}	SARS-CoV-2, influenza, tuberculosis
J_{mut}	Influenza, SARS-CoV-2 variants
J_{age}	COVID-19
J_{sdi}	HCV (male preponderance, India)
J_{mob}	All pandemic pathogens
J_{imm}	SARS-CoV-2 Omicron, influenza B

Two conditions limit transfer. First, the continuous-wave approximation of Part Two is restricted to mean durations of six months or more (R_9); that includes chronic HCV, HBV, HIV and latent tuberculosis but excludes dengue, influenza, cholera, Nipah, Japanese encephalitis, Zika, chikungunya and acute COVID-19, for which no discrete formulation is given. Second, Equations 4 and 8 carry the defects in §4.5, and these must be repaired before an intervention scenario is forecast for any disease.

7. Machine learning and AI: a specification, not a performance claim

No model was trained, because there is no data. What can be specified, and what the simulations of §4 say it must achieve, is the following.

7.1 Model structure

A hierarchical Bayesian model with partial pooling across districts is the appropriate learner at the scale of one state: tens of districts, sparse counts, and a small number of structured covariates. Counts y_{kd} of RNA-positive cases among n_{kd} screened in district k and stratum d follow a binomial distribution with prevalence p_{kd} . The log-prevalence is the sum of (i) an offset from Equations 1 and 8, (ii) covariates from the flux terms of Equation 3 (parenteral-exposure and unsafe-injection proxies, out-migration, sex ratio), (iii) stock terms of Equation 4 taken from NVHCP programme records (screening, treatment coverage, registry completeness), and (iv) a spatially structured random effect. The framework prior of §3.2 enters through the intercept, tempered as in §4.3. Deep sequence models become relevant only if facility-level time series from the programme management information system exist at scale; at the present data scale they would overfit.

7.2 What "improving the band" means

The 4.53-point planning band is the sweep of a single unmeasured parameter. Data narrow it directly: a series of 3,076 consecutive attendees gives a full interval width of 2.0 points, 56% narrower than the band. Ancillary data (blood-bank reactivity, programme testing counts, hospital registers) can pull the incidence ratio toward a value before a dedicated screen, but only if their sampling frame is modelled, and a gastroenterology OPD is enriched for liver disease. Equation 4's registry and screening terms are duration corrections; none of them corrects a biased frame.

7.3 The skill threshold

Figure 4 sets the bar. At $\sigma = 0.5$, a predictor correlation of $\rho = 0.3$ gives a proportional-allocation gain of only 1.07 times the uniform yield. We therefore propose that an ML layer be permitted to alter an allocation only when its out-of-sample district-level correlation, estimated by spatial block cross-validation, has a lower 95% confidence bound above 0.3 and a point estimate of at least 0.5. This is a proposed convention, stated in advance, not a derived constant.

8. Pre-declared validation register

Test	What is measured	Passes if	Gates
F6	Consecutive, unselected gastro-OPD series at BHU, RNA-confirmed	Prevalence in 9.06–13.60%, and Varanasi:Lucknow ratio in the same referral frame ≈ 0.977 with Varanasi below Lucknow	Use of the prior at nominal weight
F7	Posterior interval under the framework prior vs flat prior after the first 500 attendees	Intervals overlap and the flat-prior point estimate lies inside the framework interval	Choice of tempered or nominal weight
F8	District-level out-of-sample correlation ρ by spatial block cross-validation	Lower 95% bound > 0.3 and point estimate ≥ 0.5	Any change in allocation by an ML layer
F9	Stepped-wedge pilot: yield per screen in model-prioritised vs uniformly allocated districts	Ratio ≥ 1.10 with the interval excluding 1	Scale-up of targeted case-finding

An observed Varanasi:Lucknow ratio above 1.0 in F6 would refute the direction of the envelope for prevalence at this latitude pair, independently of either absolute value.

9. Limitations

1. There are no primary data. The prediction is a pre-registered expectation and the three simulations are conditional on their stated assumptions.

2. The anchor's denominator is unstated, so the population frame of the prediction is inherited, not defined. The framework has no term that corrects a biased sampling frame.
3. The band-to-prior conversion (± 2 standard deviations) is a convention. A different convention changes every figure in §4.3.
4. The targeting simulation assumes log-normal district heterogeneity, equal screening cost across districts, independent screens and no saturation of the case pool in a district. It uses 28 hypothetical districts and says nothing about the true heterogeneity of Eastern Uttar Pradesh, which is exactly the quantity that §5.3 says must be measured first.
5. The framework's own audit registers list unresolved defects (R3 to R10), and IS-1a to IS-1c are reported here. Equations 3 and 5 were used structurally; their flux and stock terms were not evaluated.
6. Equation 10's 0.40 threshold is the framework's criterion and is half the WHO service target; it is not the programme's elimination criterion.
7. Targeting changes where active case-finding is directed; it does not change who is entitled to treatment.
8. Programme facts were taken from the sources cited; the all-India community figure of 0.85% and the Lucknow anchor of 11.6% are taken from the corpus and were not independently verified here.

10. Conclusion

The twelve-equation framework is a useful *starting prior and a specification*, and a poor district map. As a prior it is worth about 782 screened attendees, which is most of the sample needed for planning-grade precision and a fifth of that needed for ± 1 pp; it must be tempered until a sentinel screen shows that it holds. As a map, its geographic term moves the Eastern Uttar Pradesh prediction by 0.16 percentage points and could add at most 0.3% to screening yield. The value of a machine-learning layer lies in populating the flux and stock terms with data and it should be adopted only if it reaches a stated skill ($\rho \geq 0.5$) in out-of-sample testing. The first step is not a model. It is one consecutive screen at one clinic.

Declarations

Data availability. No primary data. All computations are reproduced by `paper_calc.py` (random seed 20261012) and tabulated in Appendix A.

Ethics. No human subjects. The F6 screen would require institutional ethics approval and informed consent.

Competing interests and funding. [To be completed by the author.]

AI assistance. [To be completed by the author: drafting and computation in this work were AI-assisted; all quantities were regenerated from the supplied script.]

References

1. Ministry of Health and Family Welfare, Government of India. National Viral Hepatitis Control Programme (NVHCP). <https://nvhcp.mohfw.gov.in/>
2. Press Information Bureau, Government of India. India rolls out National Viral Hepatitis Control Program (PRID 1566140). <https://www.pib.gov.in/PressReleaselframePage.aspx?PRID=1566140>
3. World Health Organization. Global Health Sector Strategy on Viral Hepatitis, Annex 1: global targets. NCBI Bookshelf NBK442270.
4. Hutin YJ-F, Bulterys M, Hirnschall GO. How far are we from viral hepatitis elimination service coverage targets? J Int AIDS Soc. 2018;21:e25050. doi:10.1002/jia2.25050

5. Linkage to care of people with chronic hepatitis B and hepatitis C virus infection among blood donors: experience from an apex treatment centre under the National Viral Hepatitis Control Program, India. PMC11664039.
6. Progressing towards global hepatitis C elimination: a systematic review and meta-analysis of care cascades in key populations. Lancet Reg Health West Pac. PIIS2666-6065(26)00007-6.
7. The Synthesis of Infectious Diseases, Part Two. Author's manuscript (147 pp.), Chapters 4 to 12 and 15.
8. Hepatitis C: An Integrated Monograph, Chapters 19 to 21 (computational audit of the first edition of Part Two).
9. In-Silico Finding IS-1: predicted hepatitis C prevalence at the Model Treatment Centre, BHU, Varanasi. Companion document.

Appendix A. Computation register

Quantity	Value	Source in paper_calc.py
Transport factor $\cos 28^\circ / \cos 25.3176^\circ$	0.9768	TRANSPORT
Point prediction	11.33%	P_POINT
Prior	Beta(88.6, 693.0), effective n 782	band_prior
Sample sizes (Table 4)	exact integer bisection	n_for_halfwidth, hw_at
Coverage (Figure 3b)	1,500 replicates $\times$ 7 truths, n = 1,000	coverage
Targeting (Figure 4)	K = 28, 4,000 replicates, $\sigma \in \{0.25, 0.5, 0.75\}$	targeting
Envelope skill ceiling	$\text{sd}(\ln \cos \lambda) = 0.0041$	envelope_skill
Planning arithmetic (Tables 2, 3)	$n \times p, 1/p$	planning