WORKING PAPER WP-006 · CRITERICA INTELLIGENCE RESEARCH · NOVEMBER 2026

How to Read a Criterica Estimate

Criterica Intelligence Research

Corresponding: Jack Cosentino, founder · research@criterica.com

Published 8 November 2026 Updated 8 November 2026 Version 1.0

11-minute read

ABSTRACT

A Criterica estimate is a set of labelled figures, each with a basis, a horizon, a date and a status. This paper teaches a reader to use them well: what a band is and why it replaces a point, where dollar figures and returns come from, why direction of outcome is read only from sources that record it, what a thin segment means, how a figure responds when the law or the court changes, and what evidence turns a forecast into a claim about realized outcomes for an asset class.

IN ONE PARAGRAPH

A Criterica estimate is a set of labelled figures, each with a measure, population, horizon, date, version and status. Size on the conservative edge of a band, plan on the median and stress on the tail. Dollar figures are labelled distributions read from comparable closed matters, and returns are arithmetic on those bands and the terms you state. Direction of outcome is read only from sources that record it.

SERIES
Working Paper WP-006

VERSION
1.0

PUBLISHED
8 November 2026

STATUS
Published

AUTHORS

Criterica Intelligence Research

Corresponding author: Jack Cosentino, founder

WHO THIS IS FOR AND THE DECISION IT INFORMS

Reader. Everyone who receives a Criterica estimate: funders, buyers, lenders, insurers, law firms, general counsel, platforms and courts.

Decision. How to use each figure in an estimate well: what to size on, what to plan on, what to stress and what to ask for next.

Cost of getting it wrong. Reading a band as a point, or a computed return as a forecast, turns a sound estimate into an overconfident decision.

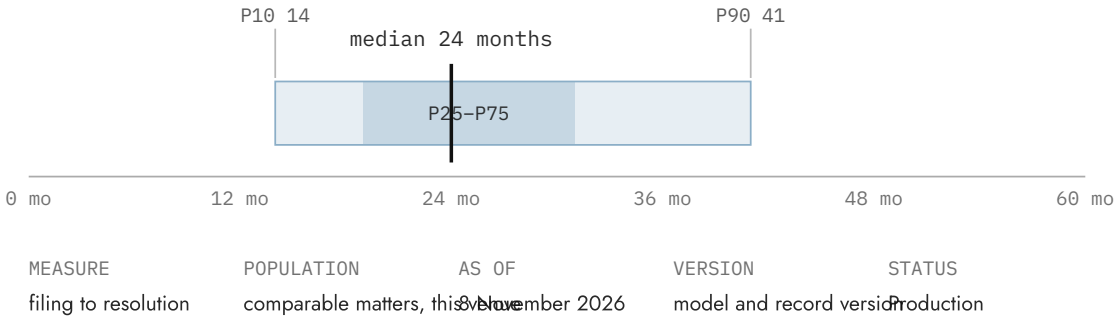
01

An estimate is a set of labelled figures

A reader who receives a Criterica estimate is usually making one of a handful of decisions: how much to advance or pay, how much to reserve, when cash will return, whether to settle now or continue, which matters in a book need attention first. Each of those decisions uses a different figure, and each figure carries the labels a reader needs to use it correctly.

This paper is a reading guide. It explains what each figure is, how it is produced and how to act on it. The labels are not small print. They are what makes the number usable: a duration median without its population is a guess, and a settlement band without its date may describe a world that has since changed.

FIGURE 1 **Anatomy of an estimate** Illustrative



A duration estimate for one fictional matter. The outer band spans P10 to P90, the inner band P25 to P75, the vertical mark is the median. The labels beneath are part of the figure. This figure is evidence of what a labelled estimate contains, on an illustrative matter.

LABEL	WHAT IT TELLS YOU
Measure	What is estimated, with its unit: months from filing to resolution, probability of resolution before trial, settlement range in dollars.
Population	The comparable matters the estimate is read against, for example this case type in this venue at this stage.
Horizon	The period the figure covers, where the measure has one.
As of	The date of the record the estimate was produced from.
Version	The model version and record version. Together with the file they regenerate the figure exactly.
Status	Production when the model has cleared its validation gates; otherwise stated, with what the reader receives instead.

02

Bands, and how to act on their edges

Criterica states estimates as bands. A duration estimate gives a median with P25 to P75 and P10 to P90 spans; a settlement estimate gives a range with its median.[1] The band is the answer, not a margin of error around it. Legal outcomes are uncertain in a way no method removes, and a point estimate hides that uncertainty instead of measuring it.

Each part of the band supports a different action, and the most useful habit is to match the action to the part:

- **Size on the conservative edge.** An advance, a purchase price or a facility limit should hold up if the matter lands near the edge that hurts you: the low end of a value band, the long end of a duration band.
- **Plan on the median.** Staffing, budget and expected cash timing for a book are planned around the middle of the distribution, with the spread carried as a range rather than a date.
- **Stress on the tail.** Reserves, liquidity and covenants are tested against P90 duration and the low tail of value, because a portfolio meets its tail matters every year.

A band also says how much a reader knows. A narrow band means comparable matters resolved consistently; a wide one means they did not, or that the file left key facts open. The width is information to act on: a wide band is the reason to request the document that would narrow it. How bands are calibrated and tested is set out in TN-003.

> Size on the edge that hurts, plan on the middle, stress on the tail.

03

Where dollar figures and returns come from

Dollar figures in a Criterica estimate are labelled distributions: a settlement band or an award band read from comparable resolved matters, adjusted for the facts in the file. Each input that feeds them carries its source, so a reader can see which amounts were stated by the seller, which are

supported by a document and which come from the record. A seller-stated medical total and a billed amount shown on an invoice are both used, and they are labelled differently.

Returns are computed, not predicted. An IRR or a multiple on invested capital in a Criterica report is arithmetic: the value distribution and the duration distribution for the matter, combined with the terms the reader states, such as the amount advanced, the fee or return structure and the waterfall. Change the terms and the computed return changes; the estimates beneath it do not.

This separation is deliberate and it protects the reader. The models estimate what the record can support: how long comparable matters take, how they resolve and the range in which they settle or are awarded. They do not estimate the realized dollars a specific matter will return to a specific investor, because that depends on terms, collection and events the court record does not contain. A return figure is therefore always shown with the terms it was computed from, so it can be recomputed by hand. WP-003 works the arithmetic of time and returns in full.

04

Reading the direction of an outcome

Readers often want to know which side is favoured. Criterica answers that question only from sources that record it, and states which source was used.

The largest public source of federal civil outcomes records how and when each case ended: settled, dismissed, decided on motion, tried. It is an excellent source for duration and for resolution path, and Criterica uses it extensively for both. Its disposition codes describe the procedural ending, not who prevailed,[2] so they are never used to state a party-prevails direction.

Where direction matters, it comes from records that carry it: judgments and orders that state the prevailing party, appellate dispositions, agency and tribunal decisions with recorded outcomes, and verdict data. A reader will therefore see resolution path and duration for most federal matters, and a direction-of-outcome figure where a source supports one, with that source named.

05

Status and thin segments

Every model carries a status. A production model has cleared its validation gates on matters it never saw during training: discrimination of at least 0.70 AUC and expected calibration error of at most 0.05 on a later time period. Its figures are issued as estimates.

Some segments are thin: a rare case type in a small venue, or a new statute with few resolved matters. A model for a thin segment that has not yet cleared its gates is held at stub status. For the reader, a stub status means two practical things. The estimate for that exact segment is not issued as a production figure. The reader receives the nearest validated comparison instead, a broader venue or case family, labelled as such, so the decision can still be informed and the reader knows which comparison they are using.

Status is reviewed as data accumulates. A stub that later clears its gates is promoted and recorded in the public model registry; a production model that stops clearing them is demoted on the record. MC-002 shows what each model card states.

06

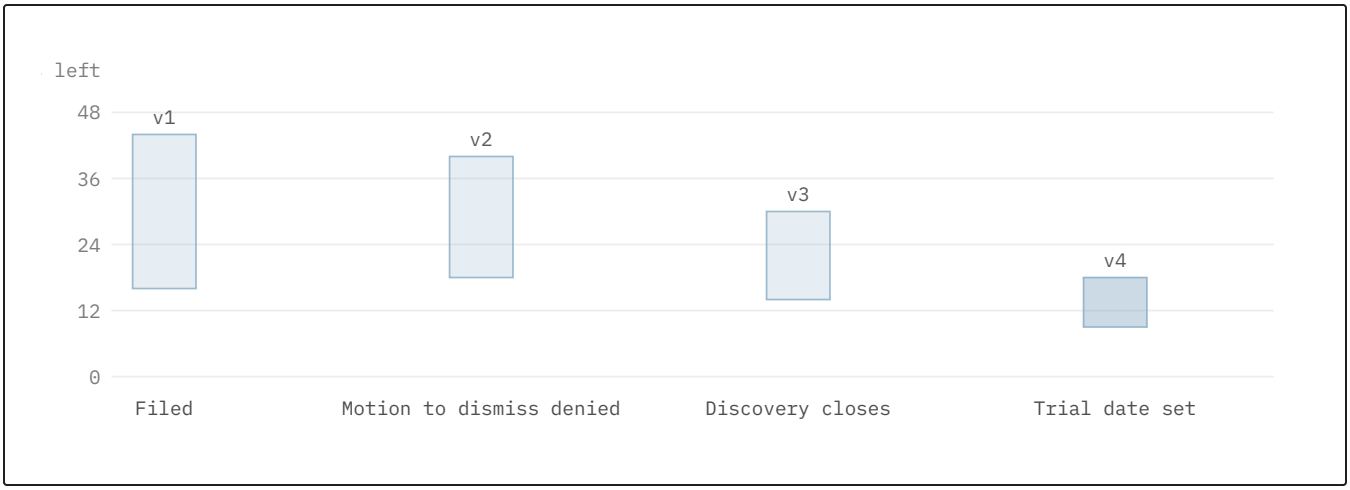
When the law or the court changes

Outcomes depend on the rules that govern them. Tort reform, a new procedural rule, a change in a court's bench or a mass consolidation of cases can shift duration and value for matters filed after it. Criterica calls these regime shifts and handles them in three ways.

1. Models are validated on later time periods, never on random samples, so a model that cannot survive a change in era does not reach production.
2. Known shifts are marked by date, so matters filed under the new regime are read against comparable matters from the same regime as soon as enough exist.

3. When a shift is too recent for the record to show its effect, the estimate says so in its labels, and the band is read with that in mind.

FIGURE 2 **An estimate changes only when the record changes** Illustrative



The duration band for one fictional matter, re-issued at four events in its record. Each version is preserved, so the change from v1 to v4 is a fact on file rather than a recollection. This figure is evidence that an estimate moves only when the record does.

Between events, the estimate does not move. A reader who asks again tomorrow receives the same figure, for the reasons set out in WP-005. When the record changes, the estimate is re-issued as a new version and the earlier one is kept, so a book's history of estimates can be compared against what eventually happened.

07

From forecast to realized-outcome claim

The current fleet estimates what the court record supports. Realized-outcome claims, statements about the dollars a class of assets actually returns, are made one asset class at a time and only with evidence. The evidence required is specific.

1. A tape of resolved matters for that asset class, supplied by counterparties, with the realized dollars recorded.

2. A backtest whose design, metrics and pass criteria are registered before the outcomes are examined.
3. A model that passes that preregistered test on matters it never saw.

When an asset class clears that bar, the claim is published with its evidence: the test design, the sample and the result. Until then, readers receive bands from the record and returns computed from their stated terms, which is exactly what the evidence supports. The preregistration design is public at /score/preregistration.

08

A reader's checklist

1. Name the decision, then pick the figure that serves it.
2. Check the population, the horizon and the as-of date.
3. Size on the conservative edge, plan on the median, stress on the tail.
4. Read any return as arithmetic on the bands and your stated terms.
5. Note the status; where a stub applies, note the comparison you received instead.
6. Ask which document would narrow the band, and request it.
7. Keep the version. A figure with its version can be regenerated and compared with the outcome later.

For the treatment of incomplete files, see WP-004. For the full methodology, see how it works.

09

The question, as the reader has it

Every reader of an estimate is deciding how much of a number to trust, and for what. An underwriter wants to know which edge to size on. A claims team wants a reserve it can defend. A general counsel wants to know when cash is likely to move. A researcher wants to know what the figure is a statement about. A point estimate serves none of them, because each needs the spread: the underwriter the adverse edge, the claims team the tail, the counsel the timing range, the researcher the population. Reading guides exist because a figure without its labels is easy to misuse and a figure with them is easy to act on.

10

What is measured

QUANTITY	UNIT	POPULATION IT IS READ AGAINST
Time to resolution	Months from filing to a recorded termination	Closed matters of the same case type, venue and stage, from filed court records
Settlement or award band	US dollars, a range with its median	Closed matters with a recorded amount, adjusted for the facts in the file
Probability of an event	A number from 0 to 1 over a stated horizon, for example resolution before trial	Closed matters that could have had the event

The inclusion logic is the same for all three: a matter is part of a comparison population only when its record shows a termination, so a case still open is never counted as an outcome and a case whose ending the record does not show is never treated as a zero. The populations are described in public terms, such as US federal civil court records or filed state court records. Where an estimate rests on a different population, the label says so.

11

How a number travels from the record to the band

1. **Record to structured event.** Filings, orders and stated amounts are read into dated events. Parties, counsel and judges are matched across records by entity resolution on a knowledge graph, so one actor is one node however the record spells the name.
2. **Structured event to model class.** Time to resolution is modelled with survival analysis (Kaplan-Meier estimation and proportional-hazards framing), which uses open matters correctly instead of dropping them. Event probabilities use calibrated probabilistic classification. Sparse cells borrow strength from their parent group through hierarchical partial pooling, in proportion to how little data they hold.
3. **Model class to calibrated output.** Raw model scores are recalibrated on held-out matters so that a stated probability matches an observed frequency, and bands are fitted with quantile and conformal methods so that their stated coverage can be tested.
4. **Output to band shown.** The reader receives the band with its labels: measure, population, horizon, as-of date, versions and status. P10, P25, P50, P75 and P90 are read directly from the fitted distribution.

Each class fits its problem for a stated reason. Survival analysis fits because the question is a time to an event that has not always happened yet. Quantile and conformal intervals fit because legal outcomes are skewed and a band is more honest than a mean. Partial pooling fits because venues are uneven, and a rule that treats a thin venue like a thick one is wrong in one of the two directions.

How the labels are earned: validation design

A figure reaches production status only after it is scored on matters it never saw. What is frozen is the model at a time cut; what is compared is each stated probability or band with how the matter actually closed; what prevents leakage is that every input, not just the outcome, is cut at the point-in-time boundary, so nothing known only later can inform an earlier estimate.

The metric families are discrimination (how well the score ranks), calibration (read from reliability diagrams and Brier-type scores, which reward honest probabilities) and interval coverage (the

share of outcomes that land inside a stated band). A band that claims 80% should contain about 80%. TN-003 explains each.

Figure 1 is evidence of how a figure is labelled and how its band is read, drawn on an illustrative matter. Figure 2 is evidence of one property: the estimate changes when the record changes and at no other time. Neither shows measured model performance, which is stated in model cards and in TN-003.

13

Scope and precision

The models estimate what the court record supports: how long comparable matters take, how they resolve and the range in which they settle or are awarded. They do not estimate the realized dollars a specific matter will return to a specific investor, which depends on terms, collection and events the court record does not contain, and they do not state which party prevails from a procedural disposition code. Returns are arithmetic on a band and the terms the reader states.

- Data lag: the record reflects filings up to its as-of date, so a figure is re-issued as the record moves.
- Censoring: open matters are not outcomes; survival methods account for them, and the band reflects what the closed matters show.
- Venue sparsity: where few comparable matters exist, the band widens and the status reflects it.

14

What a reader can reproduce

A reader can recompute any return from a band and stated terms with a calculator, check a percentile reading against the figure, and test whether a system reproduces its own number on repetition (WP-005). A reader can reproduce the public court statistics the duration reports are

built on (DR-003). Reproducing a matter-level band requires the platform, because it depends on the record version and model version that accompany the figure.

What this number means

Every figure in a Criterica estimate states what it measures, the population it is read against, its horizon, its date, its model and record versions and its status. Read together, those labels are the scope of the number: a calibrated distribution for that measure on that date, produced the same way every time the same file is read. The illustrative matter in Figures 1 and 2 is not a real file.

REFERENCES

1. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282), 457–481. doi.org/10.1080/01621459.1958.10501452
2. Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society, Series B, 34(2), 187–220.
3. Klein, J. P., & Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censored and Truncated Data (2nd ed.). Springer.
4. Gelman, A., & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
5. Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50. doi.org/10.2307/1913643
6. Romano, Y., Patterson, E., & Candès, E. J. (2019). Conformalized quantile regression. Advances in Neural Information Processing Systems, 32. arxiv.org/abs/1905.03222
7. Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: a gentle introduction. Foundations and Trends in Machine Learning, 16(4), 494–591. arxiv.org/abs/2107.07511
8. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78(1), 1–3. doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
9. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378. doi.org/10.1198/016214506000001437
10. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. Proceedings of the 34th International Conference on Machine Learning, PMLR 70, 1321–1330. arxiv.org/abs/1706.04599

11. DeGroot, M. H., & Fienberg, S. E. (1983). The comparison and evaluation of forecasters. The Statistician, 32(1–2), 12–22. doi.org/10.2307/2987588

12. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229. doi.org/10.1145/3287560.3287596

NOTES

1. Percentile notation: P10 is the value below which 10% of comparable outcomes fall, P50 the median, P90 the value below which 90% fall.
2. The largest public record of US federal civil cases records how and at what stage each case terminated. Its disposition and judgment fields describe the procedural ending, and they are not a reliable record of which party prevailed.

Prepared by Criterica Intelligence Research with data engineering, validation and editorial review. Criterica Intelligence Research publishes numbered documents in five series; each is versioned, dated and issued as a new version when it changes, with the change recorded.

CITE THIS DOCUMENT

Criterica Intelligence Research. (2026). How to Read a Criterica Estimate (Working Paper WP-006, version 1.0). Criterica. https://critericaintelligence.com/research/how-to-read-an-estimate

```
@techreport{criterica_wp006_2026,
  author = {{Criterica Intelligence Research}},
  title = {How to Read a Criterica Estimate},
  institution = {Criterica},
  type = {Working Paper},
  number = {WP-006},
  year = {2026},
  month = nov,
  note = {Version 1.0},
  url = {https://critericaintelligence.com/research/how-to-read-an-estimate}
}
```

DOI: pending