Getting Started with AI in Transfusion Medicine and Immunohaematology

Path-iQ Getting Started Series · No. 20 of 20 · A practical primer for the practising pathologist

July 2026

1. A field with enormous potential and remarkably little deployment

Transfusion medicine presents an unusual profile. The theoretical case for AI is strong: the discipline manages a perishable, donated, finite resource against unpredictable demand, makes repeated threshold-based clinical decisions, and operates one of the most data-rich supply chains in healthcare. Prediction is central to almost everything it does.

Yet real-world implementation is close to absent. A systematic review searching for studies reporting prospective deployment with genuine workflow integration screened 1,243 records, reviewed 31 full texts, and found only three studies meeting inclusion criteria. A separate structured review covering January 2018 to April 2025 identified 565 publications on AI and machine learning in transfusion medicine, from which four exemplar studies were selected to represent distinct learning paradigms.

Read those numbers together and the picture is clear: substantial publication volume, minimal clinical deployment. For a pathologist starting out, this is the most important framing in the subspecialty. The gap is not evidence of impossibility; it is an implementation gap, and implementation is where the contribution is available.

2. The task map

Blood demand and inventory forecasting. Predicting component demand by type and by day, matching it against donation supply and shelf life. Red cells last 35–42 days; platelets last five to seven. Wastage from expiry and shortage from under-supply are both routine, and both are forecasting failures. This is the clearest and most immediately actionable application.

Donor recruitment and retention. Predicting which donors will return, who is at risk of deferral for low haemoglobin, and how to target campaigns. Donor-side modelling is well suited to standard machine learning and carries minimal clinical risk.

Transfusion requirement prediction. Predicting which surgical patients will require transfusion, enabling appropriate crossmatch ordering and pre-operative optimisation. Machine learning classifiers for predicting intraoperative transfusion in non-cardiac surgery have been assessed in the literature and represent one of the better-developed clinical applications.

Patient blood management. Identifying patients with modifiable pre-operative anaemia or iron deficiency, and supporting evidence-based transfusion thresholds. The WHO’s 2025 guidance on implementing patient blood management, with its phased implementation approach and resource-adapted toolkits, gives this work a policy framework it previously lacked.

Antibody identification. Immunohaematological antibody panel interpretation is a structured logical problem — exclusion-based reasoning across a reaction grid. It is well suited to algorithmic support, particularly for complex multiple-antibody cases and for laboratories without immediate specialist access.

Blood group phenotyping and genotyping. Automated interpretation of phenotyping results, including image-based reading of reaction patterns. Work using deep learning to interpret Rh antigen phenotyping on paper-based analytical devices points toward low-cost, decentralised testing.

Adverse event classification. Automated classification of post-transfusion adverse events from clinical records, addressing the well-known under-reporting problem in haemovigilance.

3. Your first 90 days

Weeks 1–3: quantify your wastage and your shortage events. Component expiry rates and emergency issue events over the past two years give you both the business case and the target metric. Blood bank data is unusually clean and complete.

Weeks 4–8: build or evaluate a demand forecasting model retrospectively. This is the ideal starting project: the data is entirely internal, the ground truth is historical issue records, there is no patient risk, and the outcome is measurable in units and pounds. Compare a model’s predictions against what your current ordering practice actually did.

Weeks 9–12: audit your maximum surgical blood order schedule. Most schedules are inherited, rarely revised, and poorly matched to current surgical practice and patient population. Re-deriving it from your own data — with or without machine learning — usually produces immediate improvement, and it builds the analytical foundation for more sophisticated prediction later.

4. Validation specifics

Local practice governs transfusion behaviour more than in most disciplines. Transfusion thresholds, surgical technique, cell salvage availability and prescribing culture vary substantially between institutions, so a demand model from another hospital will not transfer.

Seasonality and disruption are intrinsic. Donation and demand both vary with holidays, weather, and public events, and any forecasting model must be tested across at least a full annual cycle. Models trained on pandemic-period data are particularly distorted.

Rare phenotypes and alloimmunised patients are the highest-stakes cases and the least represented in data. Patients with sickle cell disease, thalassaemia and complex alloantibody profiles require exactly the expertise that automated systems handle worst.

Regulatory status differs from diagnostics. Blood establishment computer software is regulated, and modifications to systems governing component release fall within that framework. Confirm the regulatory position before altering any process touching component labelling, release or compatibility determination.

5. Failure modes to anticipate

Forecasting failure at the tail. Models trained on ordinary demand fail on mass casualty events, major haemorrhage and unusual surgical activity — the situations where shortage causes harm.

Antibody panel over-confidence. Exclusion-based antibody identification depends on panel cell selection and reaction strength interpretation. An algorithm confident about a partially excluded specificity is dangerous, and abstention behaviour is essential.

Threshold prediction as prescription. A model predicting that a patient will require transfusion should not be interpreted as indicating that transfusion is appropriate. Prediction of practice is not endorsement of practice, and models trained on historical prescribing will faithfully reproduce historical over-transfusion.

Equity effects in donor targeting. Recruitment models optimised for yield may systematically deprioritise donor groups, which matters when rare phenotypes needed for alloimmunised patients are more prevalent in particular ancestral populations. This is a genuine clinical risk, not only an ethical concern.

6. What to watch next

The International Society of Blood Transfusion's clinical transfusion working party has been active on AI in transfusion education, research and practice, and professional-body guidance is the development most likely to unlock implementation. Watch also for the first genuinely integrated deployments — the systematic review finding only three qualifying implementation studies means that the fourth, fifth and sixth will be disproportionately influential in establishing what good looks like. For a department with clean blood bank data and a willingness to publish, that contribution is unusually available.

Vendor due-diligence checklist

Use this before any procurement conversation. The first seven questions apply to every AI tool in pathology; the last three are specific to this subspecialty. A vendor unable or unwilling to answer them in writing has told you something important.

Applies to any pathology AI tool

What is the intended use statement, and what is explicitly excluded from it?

What is the regulatory status in my jurisdiction — cleared, CE-marked, research use only?

Can I see performance on an independent external cohort, not only internal validation?

What is the prevalence of the target condition in the validation cohort versus my practice?

What is the failure mode when the model is uncertain — does it abstain, or always answer?

What monitoring does the vendor provide for performance drift after deployment?

Who is liable if an algorithmic output contributes to a diagnostic error?

Specific to this subspecialty

Was the model trained on data from an institution with comparable transfusion thresholds and case mix?

Does it cover a full annual cycle including seasonal and holiday variation?

How does it behave for alloimmunised patients and rare phenotypes?

Ask for answers in writing, and keep them. When the tool underperforms in your laboratory — and at some point it will — the gap between what was claimed and what was validated is where the conversation starts.

7. Key references

#

Source

Note

1

Li et al., *Transfusion Medicine Reviews*, 2025;39(4):150926

565 publications; four exemplar paradigms

2

AI Implementation in Transfusion Medicine: Addressing the Challenges of Clinical Adoption, 2026

1,243 records screened; 3 implementation studies

3

Blood Transfusion, 2025;23(6):558–561

AI in transfusion medicine: paradigm shift on the horizon

4

Al-Riyami et al., *Vox Sanguinis*, 2025

ISBT Clinical Transfusion Working Party perspective

5

WHO Guidance on Implementing Patient Blood Management, 2025

Phased implementation framework

6

Park et al., *Transfus Clin Biol*, 2025;32(1):1–8

ML for predicting intraoperative transfusion

Glossary of terms used in this primer

Whole-slide image (WSI) — a digitised glass slide, typically gigapixel in scale, produced by a scanner and viewed on screen rather than under a microscope.

Foundation model — a large model pretrained on vast unlabelled image data to learn general-purpose visual features, then adapted to specific tasks with comparatively little labelled data.

Weakly supervised learning — training where only a slide-level label is available (for example “contains cancer”) rather than region-by-region annotation. Most clinical pathology AI is built this way, because exhaustive annotation is unaffordable.

AUROC — area under the receiver operating characteristic curve, a single figure summarising discrimination across all thresholds. 0.5 is chance, 1.0 is perfect. It says nothing about calibration or about performance at the threshold you would actually use.

Calibration — whether a model's stated probability matches observed frequency. A well-calibrated model that says "80% confident" is right about 80% of the time. Calibration matters more than accuracy when a clinician acts on the number.

Abstention — the model declining to answer when confidence falls below a threshold, rather than returning its best guess. Essential wherever rare entities are in play.

Batch effect — systematic differences arising from scanner, stain lot, laboratory or site rather than from biology. The leading cause of models performing worse elsewhere than where they were built.

Automation bias — the human tendency to defer to a system that is usually correct, and so to stop checking independently. A property of the workflow, not of the algorithm.

Predetermined Change Control Plan (PCCP) — a regulator-agreed specification of changes a manufacturer may make to an authorised device without a new submission. Its existence acknowledges that these systems are not static.

Shadow deployment — running an algorithm in parallel with routine practice, without acting on its output, to characterise its behaviour on your own material before it touches a live workflow.

Path-iQ — Global Pathology AI Review. Prepared for practising pathologists beginning AI adoption.