27 Model passport, study report and decision

AUTHOR
R. Heller

R. Heller · Date: 2025-11-16 · Last changed: 2026-11-09

At a glance

Guiding question: Which evidence does an independent team need for exactly this target statement?

For: project groups, doctoral researchers before a colloquium or submission, reviewers and the people who decide about the next step · **Prerequisites:** accuracy, calibration and denominators (Chapter 6), reading an AI study (Chapter 7), grouped splits and transfer (Chapter 19); for the decision sheet Chapter 21 to Chapter 25

In the course: W8 · Colloquium and decision (required); project milestones from W3 to W7 · **Data:** synthetic (the teaching case of Chapter 19, rebuilt in the chapter) · **Core reading time:** about 10 minutes

27.1 Case and decision

A doctoral researcher presents a spectral tissue classifier at the faculty's research colloquium. The draft abstract says that the method "identifies tissue class A or B with 100% accuracy". The counter-reviewer, a statistician from another group, asks three questions: 100% of what – spectra or animals? Were the test spectra taken from animals the model had already seen? And who is the claim about – further spectra of known animals, new animals of the same species, or patients? The researcher has the answers, but they are spread over a lab notebook, two scripts and a slide. The supervisor has to decide whether the abstract can be submitted on Monday and what the next experiment will be. If the sentence goes out as written, the claim is about recognising known animals; read as a statement about new animals or people, it is wrong. If the abstract is withdrawn, a useful negative result goes unreported.

27.2 Core reading

Core reading · required for W8 · about 10 minutes · ends before 'Lab'

27.2.1 One target statement, one passport

A *model passport* is a one-page record that connects a single target statement with the evidence for it: research question, data, baseline, validation, uncertainty and limits. The target statement says about whom or what the result is meant to hold – for example "spectra from animals not seen in training, same species, same device". Every other field is read against it. Model cards (Mitchell

et al. 2019) and datasheets for datasets (Gebru et al. 2021) document models and data for reuse. The passport borrows their idea of a fixed set of fields but is organised around one claim, because the same model can have a solid passport for one target statement and an empty one for another.

The key assumption of this chapter is that an entry in the passport counts as evidence only if an independent team could reproduce it from the stated data, split and code, and if its unit matches the unit of the target statement. A result per spectrum does not support a statement about animals; a test on known animals does not support a statement about new ones. This is the leakage problem of Chapter 19 seen from the reporting side: Kapoor and Narayanan found data leakage in published machine-learning studies across 17 fields and proposed model information sheets to catch it (*published*, Kapoor and Narayanan 2023).

Every field has one of three states: *filled* (with its source), *open* or *not applicable*. An open field is not a weakness to hide. It must state which information would close it – a protocol, a measurement, a second data set, a decision by a named role. A reader can then see at once what is known, what is missing and how expensive the gap is.

27.2.2 Three rules for filling in

First, every result carries its numerator, its denominator and its unit: “3/6 spectra” and “1/2 animals” are two different results, and only one of them matches a statement about new animals. Second, every entry names its source – a script and its checksum, a protocol, a table in a report – so that the counter-review can find it without asking. Third, an open field is written so that someone else could close it. “Uncertainty: open” helps nobody. “Uncertainty between animals: open; closed by a pre-specified test on further animals, with the sample size justified as in Chapter 16” tells the supervisor what the next experiment has to deliver and roughly what it will cost. The same rule applies to legal and ethical fields: “data protection: open” becomes “legal basis for the reuse of care data: open; closed by the assessment of the data protection officer, requested on a stated date”.

27.2.3 A separate decision sheet

The passport describes evidence. The *decision sheet* records what is allowed to happen on the basis of that evidence: ethics, regulation, data protection, security, the responsible roles, the conditions and the date of the next review. Keeping the two apart prevents a familiar shortcut, in which a good performance figure is read as permission to use a model. A passport can be excellent and the decision still “not yet”; a passport can be thin and the decision “run the next experiment”, which is often the right outcome of a research project.

27.2.4 Choosing the reporting path

The passport is not a publication. When the work is written up, the study type determines the reporting guideline, and the guideline in turn reveals passport fields that are still open (Table 27.1).

Table 27.1: Reporting paths by study type. The guideline is chosen by the design that produced the evidence, not by the method used. PROBAST+AI is a tool for appraising prediction model studies, not a reporting checklist. Unit: study type; no data.

Study type	Typical target statement	Reporting guideline	Appraisal of risk of bias	Chapters
Diagnostic accuracy	“Among patients referred with suspected fracture, the alert has sensitivity … against the reference standard”	STARD-AI (Sounderajah et al. 2025)	–	Chapter 7, Chapter 9
Prediction model, development or validation	“In adults admitted to this unit, the model predicts delirium within 72 hours with calibration …”	TRIPOD+AI (Collins et al. 2024)	PROBAST+AI (Moons et al. 2025)	Chapter 5, Chapter 6
Early clinical evaluation of decision support	“In a supervised pilot, the tool changes workflow and safety outcomes as follows …”	DECIDE-AI (Vasey et al. 2022)	–	Chapter 20
Animal experiment	“In new animals of this species, the measurement separates tissue classes …”	ARRIVE 2.0, Essential 10 first (Percie du Sert et al. 2020); TRIPOD+AI items for any prediction model built in the study	3Rs justification (Chapter 16)	Chapter 10, Chapter 16, Chapter 19
Simulation	“Under these assumptions, a second operating team reduces waiting time by …”	STRESS-DES, -ABS or -SD (Monks et al. 2019); ODD for agent-based models (Grimm et al. 2020)	verification and validation of the model (Chapter 18)	Chapter 15, Chapter 18

Two practical consequences follow. A study can need two guidelines – an animal experiment that builds a classifier reports the animal side with ARRIVE and the model side with the relevant TRIPOD+AI items. And a guideline describes what to report, not what counts as good evidence: a complete ARRIVE checklist for a study with two test animals is complete and still uninformative.

27.2.5 The counter-review

Before the colloquium, a second group receives only the passport, the code and the synthetic or controlled data access described in it. They try to reproduce one number from a fresh R session, check whether the unit of each result matches the target statement, and write at most three findings. The authors answer each finding with a change, a reason or a new open field. This is the format of W8; it also works as an internal review before submission. The checklists in Methods in R: Reporting templates help the second group to look for missing items quickly.

Question for the workshop: Write the target statement of your project in one sentence. Which single field of your passport would change if the target statement were moved one step further – from known to new individuals, from one centre to another, or from animals to patients?

27.3 Lab

The Lab fills a passport for the teaching case of Chapter 19: four synthetic animals with three spectra each, fictitious tissue classes A and B, and a deliberately trivial reference rule. The rule returns the class already seen for a known animal; for an animal it has never seen, it returns the pre-specified default class A. All numbers below come from that fixed design.

```
spectra <- data.frame(
  animal   = rep(c("T1", "T2", "T3", "T4"), each = 3),   # four fictitious animals
  spectrum = rep(1:3, times = 4),                          # three spectra per animal
  class    = rep(c("A", "A", "B", "B"), each = 3)          # class is a property of the
        animal
)

# Reference rule: class of a known animal, otherwise the pre-specified default class A
reference_rule <- function(train, test, default = "A") {
  pred <- train$class[match(test$animal, train$animal)]
  pred[is.na(pred)] <- default
  pred
}

# Plan 1: split by spectrum (spectra 1-2 of every animal for training, spectrum 3 for
        testing)
train1 <- subset(spectra, spectrum %in% 1:2)
test1  <- subset(spectra, spectrum == 3)
correct1 <- sum(reference_rule(train1, test1) == test1$class)

# Plan 2: split by animal (T1 and T3 for training; T2 and T4 are new)
train2 <- subset(spectra, animal %in% c("T1", "T3"))
test2  <- subset(spectra, animal %in% c("T2", "T4"))
hit2 <- reference_rule(train2, test2) == test2$class
correct2 <- sum(hit2)
animal_correct <- tapply(hit2, test2$animal, all)          # an animal counts if all its
        spectra are correct

overlap1 <- intersect(unique(train1$animal), unique(test1$animal))
overlap2 <- intersect(unique(train2$animal), unique(test2$animal))

# Target values of the teaching case
stopifnot(correct1 == 4, nrow(test1) == 4,
          correct2 == 3, nrow(test2) == 6,
          sum(animal_correct) == 1, length(animal_correct) == 2,
          length(overlap1) == 4, length(overlap2) == 0)
```

The split by spectrum gives 4/4; every test animal was also a training animal. The split by animal gives 3/6 spectra and 1/2 animals with an empty ID overlap. Uncertainty follows from the denominators. The exact (Clopper–Pearson) interval for a proportion needs only `qbeta`.

```
exact_ci <- function(x, n, level = 0.95) {
  a <- (1 - level) / 2
  c(lower = if (x == 0) 0 else qbeta(a, x, n - x + 1),
    upper = if (x == n) 1 else qbeta(1 - a, x + 1, n - x))
}
results <- data.frame(
  test = c("Plan 1: known animals, per spectrum", "Plan 2: new animals, per spectrum",
           "Plan 2: new animals, per animal"),
  correct = c(correct1, correct2, sum(animal_correct)),
  n = c(nrow(test1), nrow(test2), length(animal_correct))
```

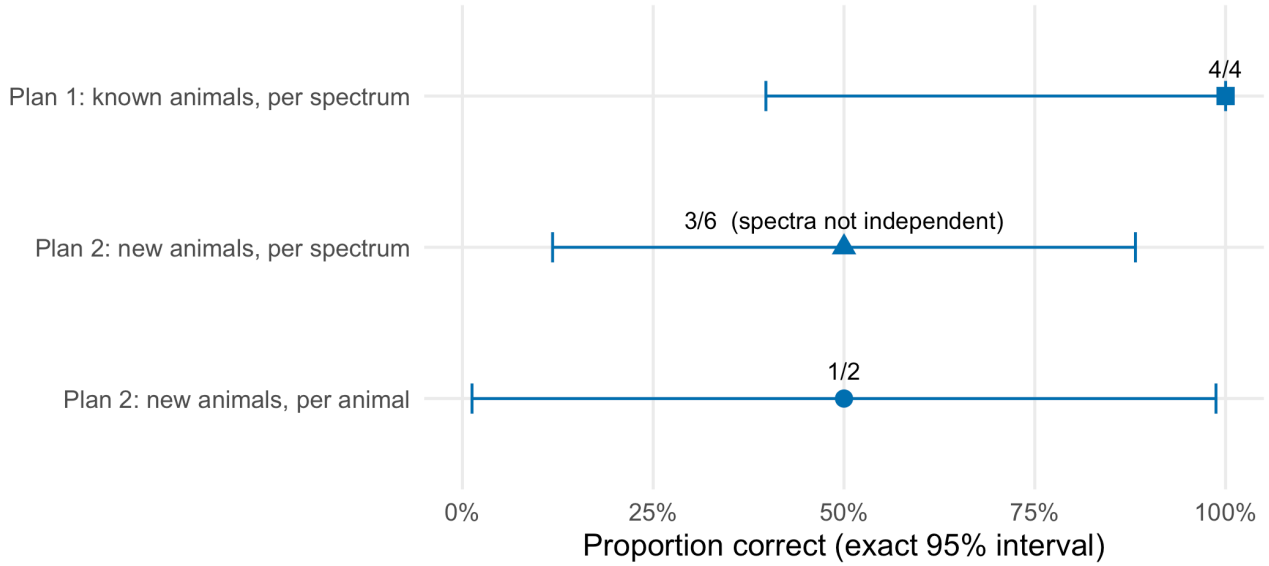
```
)
results$accuracy <- results$correct / results$n
results <- cbind(results, t(mapply(exact_ci, results$correct, results$n)))
results
```

```
#>                                test correct n accuracy      lower     upper
#> 1 Plan 1: known animals, per spectrum       4 4      1.0 0.39763536 1.0000000
#> 2   Plan 2: new animals, per spectrum       3 6      0.5 0.11811725 0.8818828
#> 3     Plan 2: new animals, per animal       1 2      0.5 0.01257912 0.9874209
```

```
stopifnot(abs(results$lower[3] - 0.0126) < 0.0005, abs(results$upper[3] - 0.9874) < 0.0005,
          abs(results$lower[1] - 0.3976) < 0.0005)
```

```
library(ggplot2)

results$test <- factor(results$test, levels = rev(results$test))
results$label <- paste0(results$correct, "/", results$n,
                        c("", "  (spectra not independent)", ""))
ggplot(results, aes(x = accuracy, y = test, shape = test)) +
  geom_errorbar(aes(xmin = lower, xmax = upper), width = 0.2, colour =
        cdsr_colours[["blue"]]) +
  geom_point(size = 3, colour = cdsr_colours[["blue"]]) +
  geom_text(aes(label = label), vjust = -1.1, size = 3.3) +
  scale_shape_manual(values = cdsr_shapes[1:3], guide = "none") +
  scale_x_continuous(limits = c(0, 1), labels = function(x) paste0(100 * x, "%")) +
  labs(x = "Proportion correct (exact 95% interval)", y = NULL,
       caption = "Synthetic teaching case: 4 animals × 3 spectra; reference rule with
        default class A.") +
  theme_cdsr()
```



Synthetic teaching case: 4 animals × 3 spectra; reference rule with default class A.

Figure 27.1: Accuracy of the reference rule in the teaching case, with exact 95% intervals. Plan 1 tests spectra of animals that were also in training; plan 2 tests two new animals. The middle row treats six spectra as independent although they come from two animals, so its interval is too narrow. Unit: spectrum or animal as labelled; n = 4 synthetic animals, 12 spectra.

27.3.1 The filled passport

The passport is a table with one row per field. Each field has a status, and every open field states what would close it. The script builds it from the objects above, so the numbers cannot drift from the analysis.

```
fmt_ci <- function(i) sprintf("%s to %s", cdsr_percent(results$lower[i]),
        cdsr_percent(results$upper[i]))
passport <- data.frame(
  field = c("Target statement", "Intended use", "Unit of analysis and independent n",
        "Data",
            "Reference standard", "Baseline", "Candidate model", "Validation", "Result",
            "Uncertainty", "Counter-check", "Limits", "Code and data access", "Artefact"),
  entry = c(
    "Assign class A or B to spectra from animals not seen in training (same fictitious
        species, device and protocol)",
    "Research only; no clinical or veterinary decision",
    "Animal; 4 animals with 3 spectra each; test set of 2 new animals",
    "Synthetic teaching data (status: synthetic); class is constant within an animal",
    "",
    "Reference rule: class of a known animal, otherwise pre-specified default class A",
    "",
    "Split by animal; ID overlap between training and test: 0",
    sprintf("%d/%d spectra and %d/%d animals correct", correct2, nrow(test2),
            sum(animal_correct), length(animal_correct)),
    sprintf("Per animal %s (exact 95%%); uninformative with 2 test animals", fmt_ci(3)),
    "",
    "One species, one device; no human data; no statement about patients",
    "Code open; data synthetic and generated by the script",
    paste("Script checksum recorded at release (see the decision sheet)")
  ),
  status = c("filled", "filled", "filled", "filled", "open", "filled", "open", "filled",
        "filled",
            "filled", "open", "filled", "filled", "filled"),
  closes_gap = c("", "", "", "",
                 "Protocol of the reference method (e.g. histology), including blinding to
        the spectra",
                 "",
                 "Result of a spectral classifier on the same animal-wise split, compared
        with the reference rule",
                 "", "", "",
                 "Repeat with changed illumination reference and report the change in the
        result (see @sec-b10)",
                 "", "", "")
)
knitr::kable(passport[, c("field", "entry", "status", "closes_gap")],
             col.names = c("Field", "Entry", "Status", "Information that would close the
        gap"))
# Every open field names the information that would close it
stopifnot(all(passport$closes_gap[passport$status == "open"] != ""),
          sum(passport$status == "open") == 3,
          !any(passport$status == "filled" & passport$entry == ""))
```

Table 27.2: Filled model passport for the teaching case. Unit: field; data status: synthetic teaching case of Chapter 19.

Field	Entry	Status	Information that would close the gap
Target statement	Assign class A or B to spectra from animals not seen in training (same fictitious species, device and protocol)	filled	
Intended use	Research only; no clinical or veterinary decision	filled	
Unit of analysis and independent n	Animal; 4 animals with 3 spectra each; test set of 2 new animals	filled	
Data	Synthetic teaching data (status: synthetic); class is constant within an animal	filled	
Reference standard		open	Protocol of the reference method (e.g. histology), including blinding to the spectra
Baseline	Reference rule: class of a known animal, otherwise pre-specified default class A	filled	
Candidate model		open	Result of a spectral classifier on the same animal-wise split, compared with the reference rule
Validation	Split by animal; ID overlap between training and test: 0	filled	
Result	3/6 spectra and 1/2 animals correct	filled	

Field	Entry	Status	Information that would close the gap
Uncertainty	Per animal 1.3% to 98.7% (exact 95%); uninformative with 2 test animals	filled	
Counter-check		open	Repeat with changed illumination reference and report the change in the result (see Chapter 10)
Limits	One species, one device; no human data; no statement about patients	filled	
Code and data access	Code open; data synthetic and generated by the script	filled	
Artefact	Script checksum recorded at release (see the decision sheet)	filled	

27.3.2 The decision sheet

The decision sheet sits next to the passport, not inside it. For the teaching case it reads as follows.

Table 27.3: Decision sheet for the teaching case, kept separate from the passport. Unit: decision; data status: fictitious project.

Dimension	Question	Entry for the teaching case	Responsible role
Ethics	Who could be harmed by an overstated claim?	Future animals in a follow-up study planned on a false premise; later patients if the claim travels into a clinical study. The next experiment needs a 3Rs justification (Chapter 16).	principal investigator with the animal welfare officer
Regulation	Does the intended purpose make this a medical device or high-risk AI system?	No: research only, no diagnostic purpose. Re-assess if the purpose changes (Chapter 22).	principal investigator
Data protection	Are personal data processed?	No (animal data). Would become relevant with human measurements (Chapter 23).	–
Security	Is the evaluated artefact the one that is reported?	Checksum of the analysis script recorded; repository access limited to the project group (Chapter 24, Chapter 25).	project data manager
Responsibility	Who decides, who reviewed?	Decision: principal investigator. Counter-review: statistician of another group.	named in the project file
Decision and conditions	What is allowed now?	Abstract may report the animal-wise result (1/2 animals) and the reference rule; it must not report 4/4 as accuracy. Next step: spectral classifier on new animals with a sample size justified in advance.	principal investigator
Review	When is this sheet looked at again?	Before any submission, and after the next experiment; date set by the principal investigator.	principal investigator

27.3.3 When an alert goes wrong: who acts next

In a clinical pilot, the decision sheet has to answer one question that matters most on a bad day: after a faulty or harmful alert, who acts next? A workable order, to be adapted locally, is the following.

1. The treating clinician decides on the patient’s care now, independently of the tool, and documents the alert and what was done.
2. The clinician reports the event to the pilot lead named in the decision sheet (row “Responsibility”). In a ward pilot this is usually the clinical lead of the unit, not the developer. The pilot lead decides whether the tool stays on, runs under stricter conditions or is switched off, using the stop criterion of the monitoring plan (Chapter 20).

3. The pilot lead informs the roles the event concerns: the developer or manufacturer for analysis, the person responsible for medical device questions if the tool may be a device (Chapter 22), the data protection officer if personal data went where they should not (Chapter 23), and information security if manipulation is suspected (Chapter 25).
4. The patient is told what happened in words they can follow (Chapter 21).

This order is an *assumption* of the book, not a legal prescription. Who holds each role is decided locally and written into the decision sheet before the pilot starts; an empty “Responsibility” row is a reason not to start.

27.3.4 A method-specific short report

The study is an animal experiment that evaluates a classification rule, so the short report follows the ARRIVE Essential 10 for the animal side and borrows the TRIPOD+AI items on data splitting and performance for the model side. The text below is generated from the analysis objects.

> **Methods (short report).** Design: synthetic teaching experiment with 4 animals, 3 spectra per animal and a class label that is constant within an animal (ARRIVE items 1, 6, 8). Sample size: fixed by design; no calculation, because the experiment is didactic (item 2; a real study would justify it). Exclusions: none (item 3). Randomisation and blinding: not applicable to synthetic data (items 4–5); a real study reports both. Analysis: a reference rule that returns the known class of an animal and the pre-specified default class A for new animals; evaluation by an animal-wise split with 2 training and 2 test animals and an empty ID overlap; exact 95% intervals (item 7; TRIPOD+AI items on data splitting and performance measures).
>
> **Results (short report).** In new animals the rule classified 3 of 6 spectra and 1 of 2 animals correctly (per animal 50%, 95% interval 1.3% to 98.7%). A split by spectrum, in which every test animal had been seen in training, gave 4 of 4 spectra correct; this measures recognition of known animals and is not reported as accuracy (item 10).

27.3.5 Counter-review

The second group reads the passport and the code. Their three findings, and the answers, are part of the record.

Table 27.4: Counter-review of the teaching passport. Unit: finding.

Finding	Answer	Status
The abstract reports 4/4, which comes from plan 1 (known animals).	Abstract changed to the animal-wise result; plan 1 is reported as a recognition check only.	closed
The per-spectrum interval for plan 2 treats six spectra as independent.	Interval reported per animal; per-spectrum result kept as description.	closed
No candidate model has been evaluated; the reference rule is trivial.	Correct; recorded as an open field with the information that would close it.	open, with plan

27.3.6 The blank passport and decision sheet

Copy the two tables below into the project file. Leave no field empty: write *open* and the information that would close it, or *not applicable* with a reason.

Table 27.5: Blank model passport.

Field	Entry	Status (filled / open / not applicable)	Information that would close the gap
Target statement (who or what, which setting)			
Intended use and users			
Unit of analysis and independent n			
Data (source, period, status, access tier)			
Reference standard			
Baseline			
Candidate model			
Validation (split, time, site; ID overlap)			
Result with denominator			
Uncertainty			
Counter-check			
Limits			
Code and data access			
Artefact (version, checksum)			

Table 27.6: Blank decision sheet.

Dimension	Question	Entry	Responsible role
Ethics	Who bears the harm of a wrong decision, and who can object?		
Regulation	Which intended purpose, which role, which legal text, which open classification?		
Data protection	Which data, which legal basis, which recipients, when deleted?		
Security	Which threat, which control, which test, who reacts?		
Responsibility	Who decides, who reviewed, who can stop?		
Decision and conditions	What is allowed now, under which conditions, with which stop criterion?		
Review	When and on which trigger is the sheet reviewed?		

27.4 Counter-check

A passport can be checked mechanically for the mismatches that matter most. The function below compares the target statement with the split, the ID overlap, the unit of the headline result and the population the data come from. Before running it, predict the number of findings for three passports: the teaching passport, the original abstract (4/4 from plan 1), and the teaching passport with its target moved from new animals to patients.

```
review_passport <- function(p) {
  findings <- character(0)
  if (p$target_unit == "new animals" && p$split != "by animal")
    findings <- c(findings, "split does not create new animals")
  if (length(p$id_overlap) > 0)
    findings <- c(findings, "IDs overlap between training and test")
  if (p$target_unit %in% c("new animals", "patients") && p$headline_unit != "animal")
    findings <- c(findings, "headline counts spectra, target counts individuals")
  if (is.null(p$interval))
    findings <- c(findings, "no uncertainty interval")
  if (p$target_population != p$data_population)
    findings <- c(findings, "no data from the target population")
  findings
}

teaching <- list(target_unit = "new animals", split = "by animal", id_overlap = overlap2,
                 headline_unit = "animal", interval = exact_ci(1, 2),
                 target_population = "animals", data_population = "animals")
abstract_claim <- list(target_unit = "new animals", split = "by spectrum", id_overlap =
        overlap1,
                       headline_unit = "spectrum", interval = NULL,
                       target_population = "animals", data_population = "animals")
```

```
moved_target <- modifyList(teaching, list(target_unit = "patients", target_population =
        "patients"))

reviews <- lapply(list(teaching = teaching, abstract = abstract_claim, patients =
        moved_target), review_passport)
reviews
```

```
#> $teaching
#> character(0)
#>
#> $abstract
#> [1] "split does not create new animals"
#> [2] "IDs overlap between training and test"
#> [3] "headline counts spectra, target counts individuals"
#> [4] "no uncertainty interval"
#>
#> $patients
#> [1] "no data from the target population"
```

```
stopifnot(length(reviews$teaching) == 0, length(reviews$abstract) == 4,
          identical(reviews$patients, "no data from the target population"))
```

The teaching passport passes; the abstract fails on four counts that all come from the same decision, the split by spectrum. Moving the target to patients produces a single finding, but it is the one no amount of re-analysis can fix: the data contain no patients. Closing it requires a new study in the target population (Chapter 19), with its own passport and its own decision sheet. Note what the function does not check: whether the reference standard is sound, whether two test animals are enough, whether the classes are clinically meaningful. Those remain questions for the human counter-review.

27.5 Consequences for patients and operations

Patients and animals. In the teaching case no patient is affected yet, but the abstract is the point at which a claim starts to travel. A “100% accuracy” that later justifies a first-in-human study would expose patients to a test whose value was never shown in new individuals. On the animal side, an overstated result can trigger confirmatory experiments that a careful passport would have redirected. A clear negative passport is therefore a contribution to the 3Rs (Chapter 16).

Regulation. The passport and decision sheet are not the technical documentation required for a medical device or a high-risk AI system; they are the research record from which such documentation could later be written. The regulatory question changes with the intended purpose (Chapter 22). The decision sheet records the current purpose and names the event – a planned clinical use – that would trigger a new assessment.

Data protection and security. Here both are minor because the data are synthetic and from animals. With human data, the decision sheet refers to the legal basis, the access tier of every product (Chapter 26) and the checksum of the evaluated artefact, so that the reported model is demonstrably the tested one (Chapter 25).

Operations and responsibility. A decision sheet without names is a wish list. Each line needs a role that exists in the organisation, and the review date needs a trigger as well as a date. For a

model in a running pilot, the decision sheet is the place where the stop criterion and the person who may apply it are written down (Chapter 20).

27.6 Look up and transfer

Reviewer question: Which entry of the passport would change if the target statement were “new patients” instead of “new animals of the same species”, and which evidence would then be required?

Self-test: A team reports a fracture alert with an assumed sensitivity and specificity of 90% each, evaluated at a referral centre where 20 of 100 examined patients had a fracture. They want to write in the passport: “useful in our emergency department”, where 5 of 100 have a fracture (the scenario of Chapter 9). Which passport field is wrong, which reporting path fits the evidence they have, and what would the decision sheet say?

Answer

Annotated sources

- Sounderajah et al. (2025) – reporting guideline for diagnostic accuracy studies using AI (an erratum appeared in September 2026); the path for diagnostic claims.
- Collins et al. (2024) – reporting guideline for prediction models using regression or machine learning; also the source of the model-side items in the animal short report.
- Moons et al. (2025) – appraisal tool for quality, risk of bias and applicability of prediction model studies.
- Vasey et al. (2022) – reporting guideline for the early clinical evaluation of AI decision support.
- Percie du Sert et al. (2020) – ARRIVE 2.0 with the Essential 10, the minimum for any report of animal research.
- Monks et al. (2019) and Grimm et al. (2020) – reporting of discrete-event, agent-based and system dynamics simulations, and the ODD protocol for agent-based models.
- Mitchell et al. (2019) and Gebru et al. (2021) – model cards and datasheets, the documentation formats the passport builds on.
- Kapoor and Narayanan (2023) – survey of leakage in machine-learning-based science and the case for structured model information sheets.

27.7 Further reading

- Chapter 19 – the split plan, ID checks and external tests behind the teaching passport.
- Chapter 7 – how to read a published AI study with the same guidelines.
- Chapter 16 – sample size, small denominators and the 3Rs for the next animal experiment.
- Methods in R: Reporting templates, in particular TRIPOD-AI and ARRIVE – fill-in templates for the two guidelines of the short report.
- Methods in R: Writing a report – Reporting guidelines – where guidelines fit into a manuscript.

- CTTIR Courses: Writing a report – IMRaD, effect sizes with intervals and reproducibility in one page.
- CTTIR Courses: Pre-registration and statistical analysis plans – fixing the target statement and the analysis before the next experiment.
- CTTIR Courses: TRIPOD-AI, fairness auditing, reproducibility at scale – a lab that walks through the reporting items of a prediction model.