RAPID

Risk of Attribute Prediction-Induced Disclosure in synthetic microdata

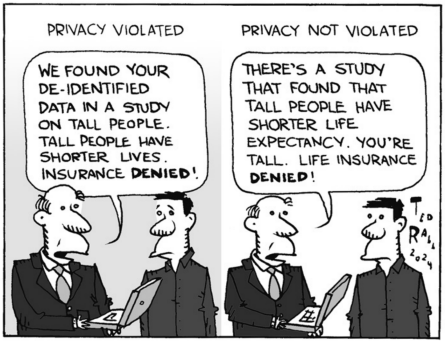
an inference-risk metric, now in the new `riskutility` package (CRAN) for risk and utility on tabular data

Oscar Thees ▪
Matthias Templ ▪
Roman Müller

FHNW — University of Applied Sciences Northwestern Switzerland

useR! 2026

Same harm, but is it a privacy violation?



Both panels end the same way, yet only one is labelled a violation. Why?

Cartoon: Ted Rall (2024), in S. L. Garfinkel, *Differential Privacy*, MIT Press.

Two answers to the same cartoon

Differential privacy: *not* violated

(Dwork, 2006)

- The promise: the outcome is *the same whether or not your record is in the data*.
- The study, and its conclusion, exist regardless of you.
- So a population-level inference is not your privacy being breached.

Predictive privacy: *still* violated

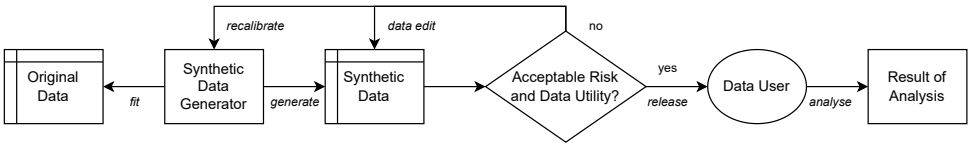
(Mühlhoff, 2021)

- The harm is being subjected to the inference, not whose data trained it.
- Models built on *others* now predict *your* sensitive traits, so an individual "opt-out" cannot protect you.
- Sibling school: group privacy (Taylor et al., 2017).

RAPID takes the second view — it measures the inference DP leaves on the table.

Synthetic microdata: a new default?

- Increasingly the go-to release mechanism: `synthpop`, `simPop`, GANs, ...
- Intuition: it *breaks the link* between real people and released records, "so no risk".
- No silver bullet: it can still re-sample *unique* real records, and any *dependencies* it keeps are what an attacker infers from. So measure, don't assume.



The custodian releases only once the answer is yes: the gate where RAPID (and `riskutility`) informs the decision.

RAPID in one sentence

Core idea

Simulate a realistic attacker as a machine-learning model: train it on the *synthetic* data to predict the sensitive attribute from quasi-identifiers, then test how well it recovers the *true* values of *real* individuals.

Inputs

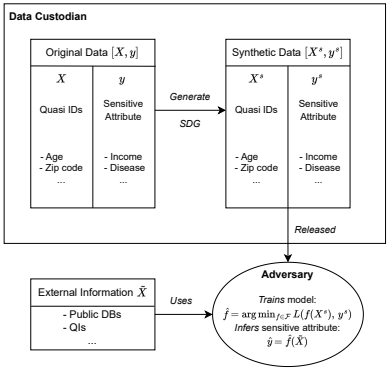
- Original data (held by custodian)
- Its synthetic version, to be released

↪ the QI columns, plus one *separate* sensitive column

Outputs

- Record-level risk flags
- A single confidence / risk rate
- Which QI patterns drive the risk

The attacker pipeline



1. **Train** on synthetic data: predict sensitive y from quasi-identifiers X (RF, CART, ...).
2. **Predict** y for each *real* record.
3. **Score** vs. truth; **flag** the confident, correct ones.

Why a model?

It infers even when *no record matches exactly*: the realistic attacker that exact-match methods miss.

Baseline correction: not all "correct" guesses are risk

- If 90% of patients are healthy, guessing "healthy" is right 90% of the time: the marginal, not disclosure. RAPID scores the attacker *relative to a baseline*, so only information beyond what anyone could already guess counts as risk.

Categorical: normalised gain

$$\text{gain}_i \;=\; \frac{g_i - b_i}{1 - b_i} \;>\; \tau$$

g_i = model's probability of the *true* class.

b_i = base rate of that class.

$\Rightarrow$ fraction of the achievable edge $1 - b_i$ won.

Continuous sensitive attribute

Symmetric relative error (default), stabilised relative, or absolute, thresholded by ε.

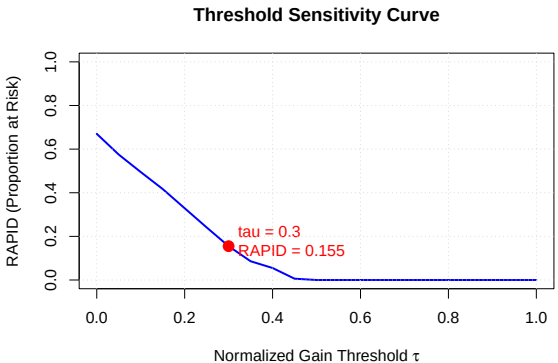
Using RAPID: one call

```r
library(RAPID); library(synthpop)
data_syn <- syn(data_orig, method = "cart")$syn

result <- rapid(
  original_data       = data_orig,
  synthetic_data      = data_syn,
  quasi_identifiers   = c("age", "education", "gender"),
  sensitive_attribute = "disease_status",
  model_type          = "rf",       # rf | cart | gbm | logit | lm
  cat_eval_method     = "RCS_marginal",
  cat_tau             = 0.3
)
print(result); summary(result); plot(result)   # familiar S3 methods
```

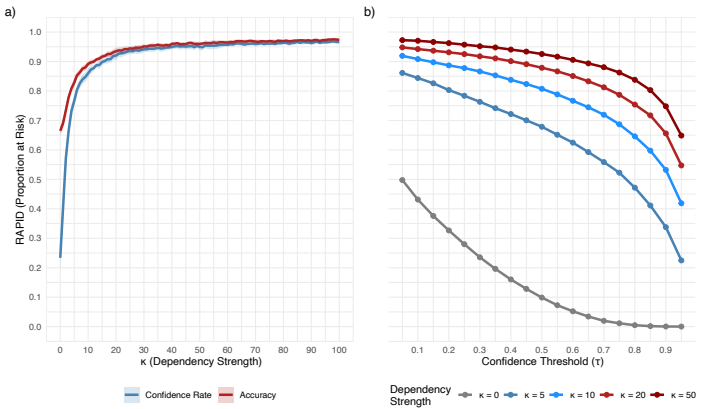
Choose your operating point

```
print(result)
#> Risk level: 15.5 %
#> Records at risk: 155 / 1000
#> Threshold (tau): 0.3

# the risk--threshold curve
plot(result)
```

Raise the threshold, fewer records are flagged. You pick where to stand.

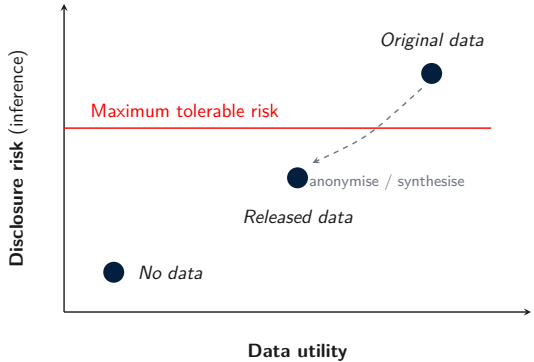


Simulation: RAPID tracks injected risk



So far: one specific risk

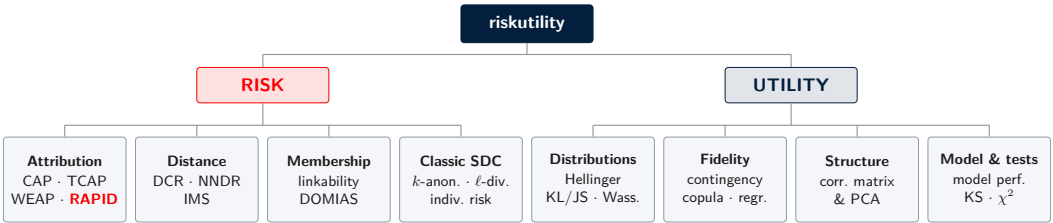
RAPID measured *one* risk. But a release is a trade-off: utility matters too.



In practice: one risk, one utility measure. Really *many* of each, so go multivariate.

`riskutility`: the metric map

Live on CRAN v0.1.0 with ∼**100 functions** across risk & utility. A selection:



`riskutility` in action: build once, measure many

```
obj <- from_synthpop(syn_obj,
  original    = orig,
  key_vars    = c("age","education","gender"),
  target_var = "disease_status")

rapid(obj, cat_eval_method = "RCS_marginal",
           cat_tau = 0.3)    # risk
pMSE(obj)                    # utility
```

Build the `synth_pair` once → then risk and utility from the same object.

Risk: `rapid(obj)`

`RAPID: 15.5%` at risk

155 / 1000 records

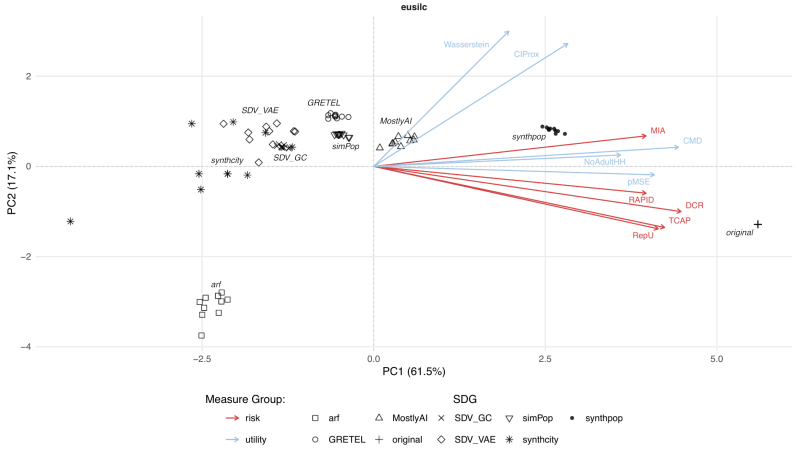
Utility: `pMSE(obj)`

`pMSE ratio: 0.81`

$\approx 1 \Rightarrow$ indistinguishable

$\Rightarrow$ one point on the risk–utility map.

Beyond the trade-off curve: the multivariate R–U map



Takeaways

- Synthetic data *breaks the link*, not inference risk (a predictive-privacy harm).
- `RAPID` measures that *one* risk: a model-based attacker with baseline-corrected scoring, record-level, categorical and continuous.
- But a release is a trade-off, and risk and utility are each multivariate.
- `riskutility` (CRAN) puts RAPID plus ∼100 risk & utility measures on one `synth_pair`: the whole R–U map, in R.

Try it

`install.packages("riskutility")` ▪ cran.r-project.org/package=riskutility

References & further reading

- C. Dwork (2006). Differential privacy. *ICALP 2006*, LNCS 4052, 1–12.
- S. L. Garfinkel (2022). *Differential Privacy*. MIT Press Essential Knowledge. (cartoon: Ted Rall, 2024)
- R. Mühlhoff (2021). Predictive privacy: towards an applied ethics of data analytics. *Ethics and Information Technology* 23, 675–690.
- L. Taylor, L. Floridi & B. van der Sloot, eds. (2017). *Group Privacy: New Challenges of Data Technologies*. Springer.
- S. Wachter & B. Mittelstadt (2019). A right to reasonable inferences. *Columbia Business Law Review* 2019(2), 494–620.
- J. Taub et al. (2018). Differential correct attribution probability for synthetic data. *PSD 2018*, LNCS 11126, 122–137.
- M. Templ & O. Thees (2026). *riskutility*. R package v0.1.0, CRAN.
- O. Thees, M. Templ & R. Müller (2026). *RAPID*. `github.com/qwertzlbry/RAPID`.

Thank you

Questions?