An odds ratio of 0.43 for a risk ratio of 0.58

Cases reporting a disability complete a referral at 26.4% against 45.5%. That is 58% as likely, and the logistic model prints 0.43. The two numbers describe the same data, only one of them is what a reader thinks they are reading, and the gap widens exactly when the outcome is common.

Pierre Robentz Cassion

Lesson 3 of 8

Regression for Programme Data — Outcomes that are yes or no

What this lesson covers

- Fit it, and read what it prints
- Why they differ, and when it gets worse
- The trap that catches analysts, not just readers
- When the odds ratio is the right number
- The sentence to write, and three not to
- Report it whole
- What comes next

Fit it, and read what it prints — In Python (cont.)

```
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf

referrals = pd.read_csv("protection-referrals-2024.v1.csv")
referrals["disability"] = referrals["disability_reported"].map(
    {"true": 1, "Yes": 1, "false": 0, "No": 0})

consenting = referrals[referrals["consent_to_refer"]].copy()
consenting["completed"] = (consenting["referral_accepted"]
                           & consenting["days_to_first_service"].notna()).astype(int)
d = consenting.dropna(subset=["disability", "case_category", "age_band",
                              "sex", "service_requested", "admin1"])

crude = smf.logit("completed ~ disability", data=d).fit(disp=0)
print(np.exp(crude.params).round(3))
```

Fit it, and read what it prints — In Python (cont.)

```
print(np.exp(crude.conf_int()).round(3))
```

Fit it, and read what it prints — In R

```r
library(dplyr)

crude <- glm(completed ~ disability, data = d, family = binomial())
exp(cbind(OR = coef(crude), confint(crude)))
```

Fit it, and read what it prints

- Odds ratio 0.430, 95% CI 0.308 to 0.601

Fit it, and read what it prints — In Python

```
rates = d.groupby("disability")["completed"].agg(["sum", "size", "mean"])
print(rates.round(4))
p1, p0 = rates.loc[1, "mean"], rates.loc[0, "mean"]
print(f"risk ratio        {p1 / p0:.3f}")
print(f"risk difference   {p1 - p0:+.4f}")
print(f"odds ratio        {(p1/(1-p1)) / (p0/(1-p0)):.3f}")
```

Fit it, and read what it prints — In R

```
d |> summarise(k = sum(completed), n = n(), rate = mean(completed), .by = disability)
```

Fit it, and read what it prints

Group	**Completed**	**n**	**Rate**
Disability reported	52	197	**26.4%**
Not reported	629	1,384	**45.5%**

Fit it, and read what it prints

Measure	Value	Reads as
Risk ratio	**0.581**	58% as likely to complete
Risk difference	**−19.1 points**	19 fewer completions per 100 cases
Odds ratio	**0.430**	—

Fit it, and read what it prints

- The model printed 0.430 and the answer is 0.581 — Both are correct; they are answers to different questions, and only…

Why they differ, and when it gets worse — In Python

```
for p0 in (0.02, 0.10, 0.25, 0.45, 0.70):
    p1 = 0.6 * p0                          # a true risk ratio of 0.6 throughout
    odds = (p1 / (1 - p1)) / (p0 / (1 - p0))
    print(f"baseline {p0:5.0%}   risk ratio 0.60   odds ratio {odds:.3f}")
```

Why they differ, and when it gets worse — In R

```
# One true risk ratio, five baselines, five different odds ratios.
```

Why they differ, and when it gets worse

Baseline risk	**True risk ratio**	**Odds ratio**
2%	0.60	0.59
10%	0.60	0.57
25%	0.60	0.53
45%	**0.60**	**0.45**
70%	0.60	0.31

Why they differ, and when it gets worse

- The odds ratio is not a fixed distortion of the risk ratio — it depends on the baseline — At a 2% outcome the two are…
- Programme outcomes are not rare — Referral completion is 43%, attendance is 88%, enrolment is 97%, and at those…

The trap that catches analysts, not just readers — In Python

```
adjusted = smf.logit(
    "completed ~ disability + case_category + age_band + sex"
    " + service_requested + admin1", data=d).fit(disp=0)
print(f"crude OR    {np.exp(crude.params['disability']):.3f}")
print(f"adjusted OR {np.exp(adjusted.params['disability']):.3f}")
```

The trap that catches analysts, not just readers — In R

```r
adjusted <- glm(completed ~ disability + case_category + age_band + sex +
                  service_requested + admin1, data = d, family = binomial())
exp(coef(adjusted))["disability"]
```

The trap that catches analysts, not just readers

	Odds ratio	**Risk difference**
Crude	0.430	−19.05 points
Adjusted	**0.388**	−**19.41 points**

The trap that catches analysts, not just readers

- The odds ratio moved by 10% and the risk difference moved by a third of a point — The usual reading — "adjustment...
- A collapsible measure equals the average of the subgroup measures — Risk differences and risk ratios are; odds ratios...
- So an adjusted odds ratio and a crude odds ratio cannot be compared to judge confounding — That comparison is the...

When the odds ratio is the right number

- A case-control study — Cases and controls are sampled separately, so risks are not estimable at all and the odds ratio. . .
- A rare outcome — Below about 10%, the odds ratio approximates the risk ratio closely enough that the distinction stops. . .
- Comparing to published literature that reports odds ratios — Then report both, and lead with the risk difference

When the odds ratio is the right number — In Python

```
def report(label, p1, p0):
    return (f"{label}: {p1:.1%} vs {p0:.1%} -- "
            f"risk difference {p1-p0:+.1%}, risk ratio {p1/p0:.2f}, "
            f"odds ratio {(p1/(1-p1))/(p0/(1-p0)):.2f}")

print(report("Referral completion, disability reported", p1, p0))
```

When the odds ratio is the right number — In R

```
# Print all three. The reader picks; you do not pick for them by omission.
```

The sentence to write, and three not to

- Not this — "Cases reporting a disability were 57% less likely to complete a referral (OR 0.43)." Two errors in one...
- Nor this — "Disability halved the odds of completion." Correct arithmetic, and "halved" will be read as risk by...
- Nor this — "OR 0.43 (95% CI 0.31–0.60, $p < 0.001$)." Complete, checkable, and it tells a programme manager nothing they...
- This — "Cases reporting a disability completed a referral at 26.4% against 45.5% for cases with no disability reported...
- The last clause is the one that does the work — An odds ratio in a programme report with no translation beside it will...

Report it whole — Example (cont.)

```
Referral completion by disability status, 1,581 consenting cases

  Disability reported     26.4%   52 of 197
  Not reported            45.5%   629 of 1,384

  Risk difference   -19.1 points   95% CI -25.7 to -12.4
  Risk ratio          0.58
  Odds ratio          0.43 (crude), 0.39 (adjusted)

  The odds ratio is reported for comparability only. Completion is a common
  outcome (43% overall), so the odds ratio overstates the relative
  difference: 0.43 in odds is 0.58 in risk.

  The adjusted odds ratio differs from the crude one partly because the odds
  ratio is not collapsible, not only because of confounding. The adjusted
  risk difference is -19.4 points, essentially unchanged from crude.
```

Report it whole — Example (cont.)

```
Fitted on the 1,581 consenting cases with complete covariates. The
statistics course reported this gap on all 1,638 consenting cases and got
-19.4 points; the 57-case difference is the complete-case restriction.
```

Report it whole

- The last paragraph is two lines and prevents the single most common over-claim — that adjustment "strengthened" a...

What comes next

- A logistic model prints coefficients on a scale nobody thinks in.

Read the full lesson, with runnable code

https://data-analysis.cassion.dev/courses/regression-for-programmes/
reg-03-the-odds-ratio-your-reader-misreads