

Thirty-three cases each

Guidance puts an active GBV caseload near 25 cases per worker. One area here averages 33.2 and peaks at 51. It also holds 1.78 case plan reviews per case against 2.45, and loses contact with 38.9% of the cases it closes. Three tables, one cause.

Pierre Robentz Cassion

Lesson 5 of 8

Protection and GBV Data — The caseload

What this lesson covers

- Caseload is a stock, not a flow
- The number the guidance is written about
- What an overloaded caseload does
- The establishment error
- What caseload is not
- Report it as a supervision table
- What comes next

Caseload is a stock, not a flow — In Python (cont.)

```
import pandas as pd

cases = pd.read_csv("protection-case-management-2024.v1.csv")
CUTOFF = 12

def months_open(row):
    opened = int(row["opened_month"][5:7])
    closed = int(row["closed_month"][5:7]) if pd.notna(row["closed_month"]) else CUTOFF
    return range(opened, max(closed, opened) + 1)

active = [
    {"caseworker_id": row["caseworker_id"], "admin2": row["admin2"], "month": month}
    for _, row in cases.iterrows()
    for month in months_open(row)
]

active = pd.DataFrame(active)
```

Caseload is a stock, not a flow — In Python (cont.)

```
caseload = active.groupby(["caseworker_id", "month"]).size().rename("open_cases")  
print(caseload.describe().round(1))
```

Caseload is a stock, not a flow — In R

```
library(dplyr)

cases |>
  mutate(opened = as.integer(substr(opened_month, 6, 7)),
         closed = coalesce(as.integer(substr(closed_month, 6, 7)), 12L)) |>
  rowwise() |>
  mutate(month = list(seq(opened, max(closed, opened)))) |>
  tidyr::unnest(month) |>
  count(caseworker_id, month, name = "open_cases")
```

Caseload is a stock, not a flow

- This is the person-time expansion from module 4's first course — applied to a caseworker rather than to a patient

The number the guidance is written about — In Python

```
by_area = caseload.reset_index().merge(  
    cases[["caseworker_id", "admin2"]].drop_duplicates("caseworker_id"),  
    on="caseworker_id",  
)  
summary = by_area.groupby("admin2")["open_cases"].agg(["mean", "max"]).round(1)  
print(summary.sort_values("mean", ascending=False))
```

The number the guidance is written about — In R

Mean and peak monthly caseload per area.

The number the guidance is written about

Area	Mean caseload	Peak
Port-de-Paix	33.2	51
Gonaives	28.2	41
Saint-Marc	24.7	43
Hinche	20.7	44
Mirebalais	18.7	32
Saint-Louis-du-Nord	14.8	26

The number the guidance is written about

- GBV case management guidance puts an active caseload at around 25 — Port-de-Paix sits above it all year and peaks at. . .
- Report the mean and the peak — A worker at 51 open cases in one month has, that month, roughly four working hours per. . .

What an overloaded caseload does — In Python

```
quality = cases.groupby("admin2").agg(  
    cases=("case_id", "size"),  
    reviews_per_case=("case_plan_reviews", "mean"),  
)  
closed = cases[cases["closure_reason"].notna()]  
quality["lost_contact"] = closed.groupby("admin2")["closure_reason"].apply(  
    lambda s: (s == "lost-contact").mean()  
)  
print(quality.round(2).sort_values("reviews_per_case"))
```

What an overloaded caseload does — In R

```
cases |>  
  summarise(n = n(), reviews = mean(case_plan_reviews),  
            lost = mean(closure_reason == "lost-contact", na.rm = TRUE),  
            .by = admin2)
```

What an overloaded caseload does

Area	Caseload	Reviews per case	Lost contact
Saint-Louis-du-Nord	14.8	1.74	26.0%
Port-de-Paix	33.2	1.78	38.9%
Mirebalais	18.7	2.10	30.8%
Gonaives	28.2	2.31	24.7%
Saint-Marc	24.7	2.38	22.1%
Hinche	20.7	2.45	20.2%

What an overloaded caseload does

- Port-de-Paix carries the highest caseload, holds among the fewest case plan reviews, and loses contact with the largest. . .

The establishment error — In Python

```
areas_per_worker = cases.groupby("caseworker_id")["admin2"].nunique()  
print(areas_per_worker[areas_per_worker > 1])
```

The establishment error — In R

```
cases |> summarise(areas = n_distinct(admin2), .by = caseworker_id) |>  
  filter(areas > 1)
```

The establishment error

- A caseload computed per worker is right. A caseload computed per worker per area is wrong — because it splits one. . .

What caseload is not

- It is not a productivity measure — A worker with 40 open cases is not working harder than one with 15; they are more. . .
- It is not comparable across case types — A child protection case with a case plan involving a school and a guardian is. . .

What caseload is not — In Python

```
mix = cases.groupby(["admin2", "case_category"]).size().unstack(fill_value=0)
print((mix.div(mix.sum(axis=1), axis=0) * 100).round(1))
```

What caseload is not — In R

```
cases |> count(admin2, case_category) |>  
  mutate(share = n / sum(n), .by = admin2)
```

What caseload is not

- Check the case mix before comparing caseloads — An area with more child protection cases is carrying more work per. . .

Report it as a supervision table — Example

Caseload, 2024, 1,108 cases across 17 caseworkers

Area	Mean	Peak	Reviews/case	Lost contact
Port-de-Paix	33.2	51	1.78	38.9%
Gonaives	28.2	41	2.31	24.7%
Saint-Marc	24.7	43	2.38	22.1%
Hinche	20.7	44	2.45	20.2%
Mirebalais	18.7	32	2.10	30.8%
Saint-Louis-du-Nord	14.8	26	1.74	26.0%

Guidance: active caseload around 25 cases per worker.

One caseworker identifier appears in two areas after a transfer; caseload is computed per worker, not per worker per area.

Port-de-Paix is above the guideline all year and shows the pattern that follows: fewest case plan reviews and most cases closed for lost contact.

Report it as a supervision table

- Put the guideline in the table — A caseload of 33 means nothing to a reader who does not know what 25 is, and the whole. . .

What comes next

- Cases close, and how long they take is the other number a supervisor is judged on.

Read the full lesson, with runnable code

`https://data-analysis.cassion.dev/courses/protection-and-gbv-data/
prot-05-caseload`