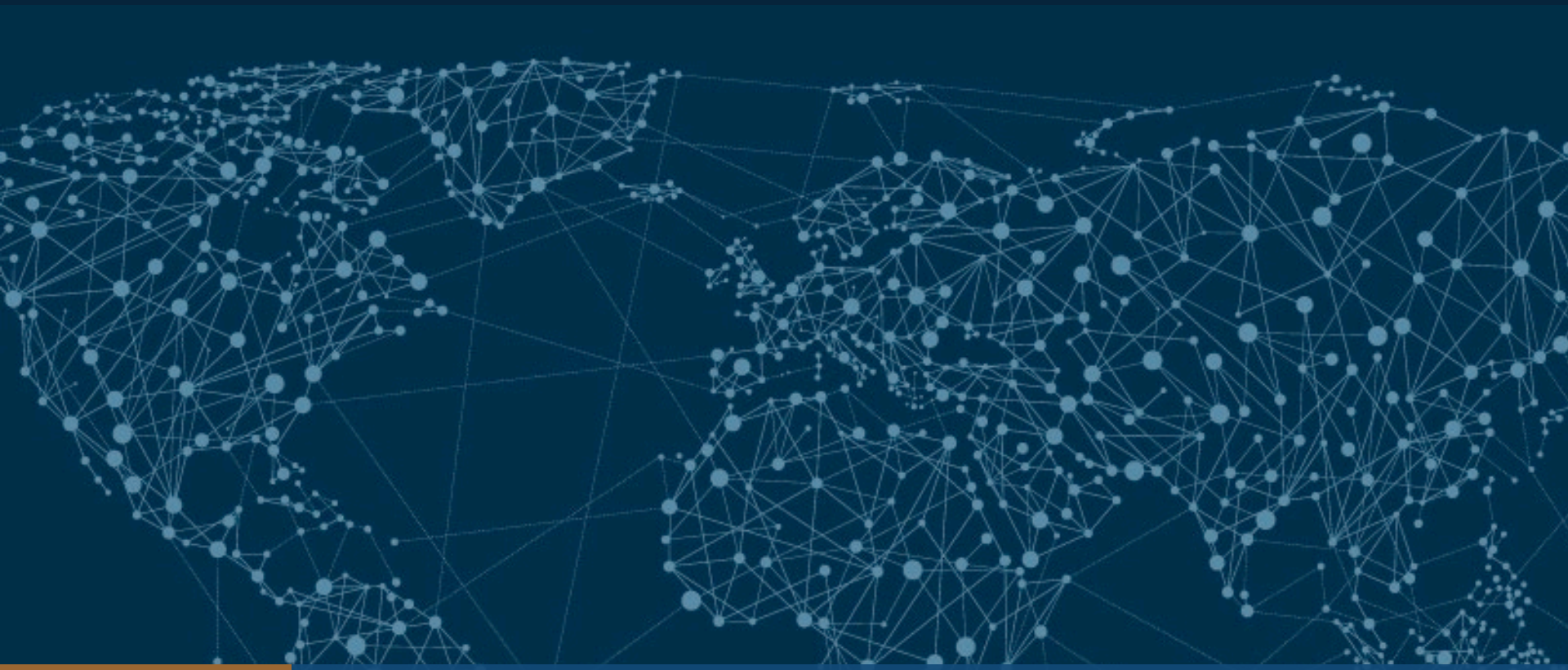


ITHI: Metrics Related to Whois Accuracy & Domain Name Abuse

ICANN DNS SYMPOSIUM, May 13th 2017

What is **ITHI**? Why Should You be Interested?

ITHI, or Identifier Technologies Health Indicators is a new ICANN initiative to "**measure**" the "**health**" of the "**identifiers**" that "**ICANN helps coordinate**".

The goal is to produce a set of indicators that will be **measured and tracked over time** that will help determine if the set of identifiers is overall doing better or worse.

This is a long term project, expected to run for many years. We are at the beginning, and it is important to get the community involved in the definition phase of the project.

What is the Scope of **ITHI**?

The scope of ITHI is all the unique Identifiers ICANN helps coordinate. It includes:

- DNS names
- IP addresses

This presentation will focus on the name track.
The RIRs are working on the number track.

At some point, both tracks will be converged.

ICANN Strategic Plan 2016-2020

https://www.icann.org/en/system/files/files/strategic-plan-2016-2020-10oct14-en.pdf

2.1 Foster and coordinate a healthy, secure, stable, and resilient identifier ecosystem.

KEY SUCCESS FACTORS (OUTCOMES)

- Increased collaboration with the global community that improves the security, stability and resiliency of the unique identifier ecosystem (including updates of the root zone, Internet numbers registries, and protocol parameter registries, operation of the “L” root server, and other operational infrastructure supporting the identifier ecosystem).
- Ecosystem is able to withstand attacks or other events without loss of confidence in the operation of the unique identifier system.
- Unquestionable, globally recognized legitimacy as coordinator of unique identifiers.
- Reduction of government/industry/other stakeholders’ concerns regarding availability of IP addresses.
- …

ITHI: Methodology

1) Define Health / Problem Areas
2) Define Metrics to measure the above
3) Get data to compute above metrics

5 Problem Areas

In the first phase of the name track of ITHI, we will focus on **5 problem areas**:

- ❑ Bad Data
- ❑ Abuse
- ❑ Excessive Traffic
- ❑ Leakage
- ❑ Lies

Over time, new problem areas could be defined, and/or some could removed.

Recap From ICANN58

Focus on Two Areas

- **Data accuracy**
  - Direct measurement by analyzing “whois” data,
  - Indirect measurement by looking at complaints received by ICANN's Compliance "department" regarding the accuracy of “whois” information.

- **Abuse**
  - Tie to Anti-abuse project from ICANN OCTO.
  - For every TLD/Registry/Registrar, calculate daily abuse score by looking at percentage of registrations that appear in various anti-abuse lists.

The Devil is in the Details

- **We need to get the detail rights about:**
  - What is measured,
  - How it is measured.

- **Plan to move forward**
  - Invite community to a series of workshops to define exactly how to do this,
  - Build a prototype,
  - Show preliminary numbers at an upcoming ICANN meeting.

Whois (In-)Accuracy

ITHI Ask to ICANN Compliance

We asked ICANN compliance department for sample data on whois inaccuracy complaints it receives to build a **prototype of a candidate metric M1**.

- We asked monthly data for 5 registrars and 5 registries covering 2016.
- The choice of registrars and registries was “random”, but covering both established and newer actors.
- Because this is only a limited sample and the methodology is still under development, we have anonymized the data to avoid singling out anybody.

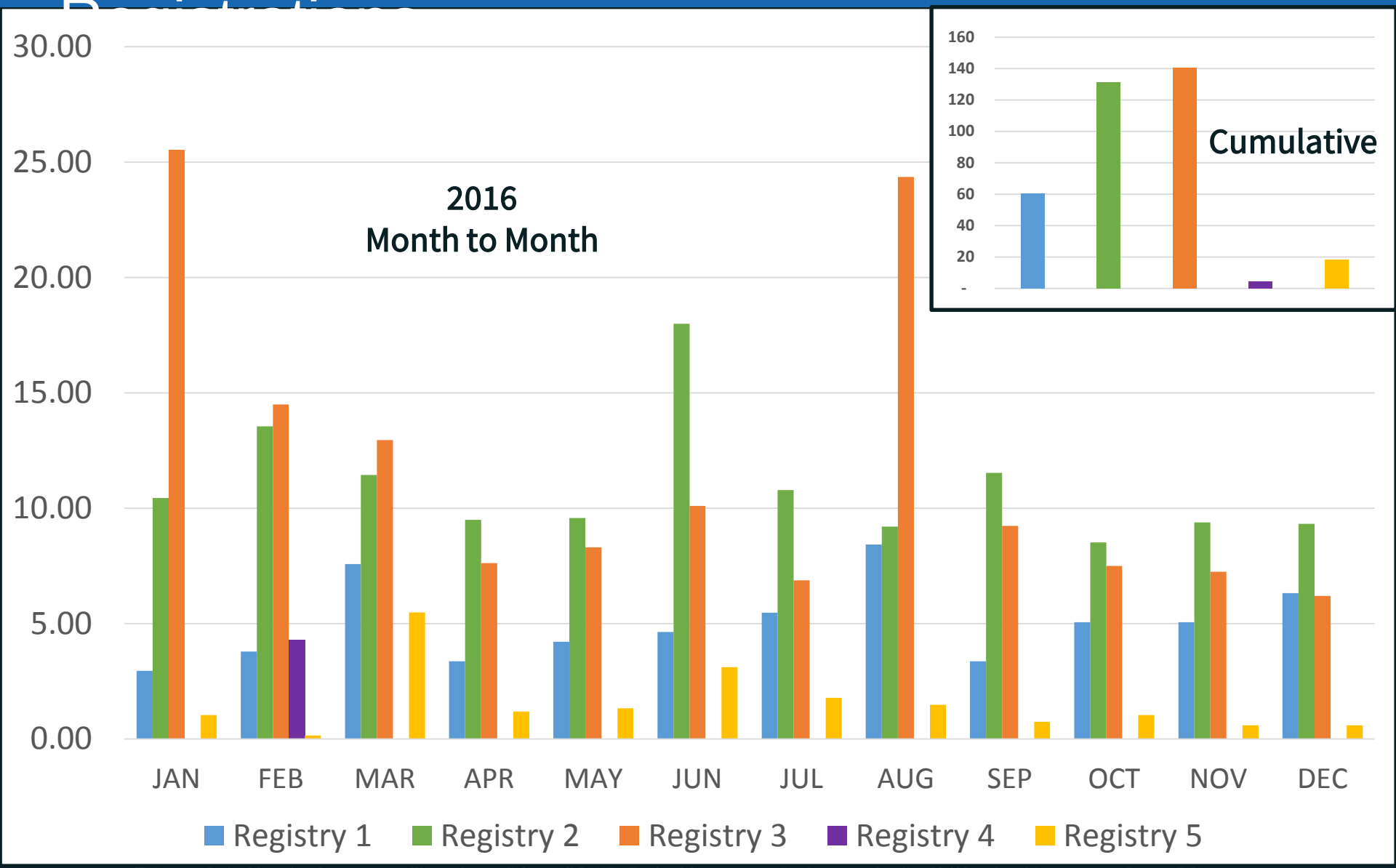
Candidate Metric Related to Data (in-)Accuracy

M1	Number of “validated complaints” per million registrations

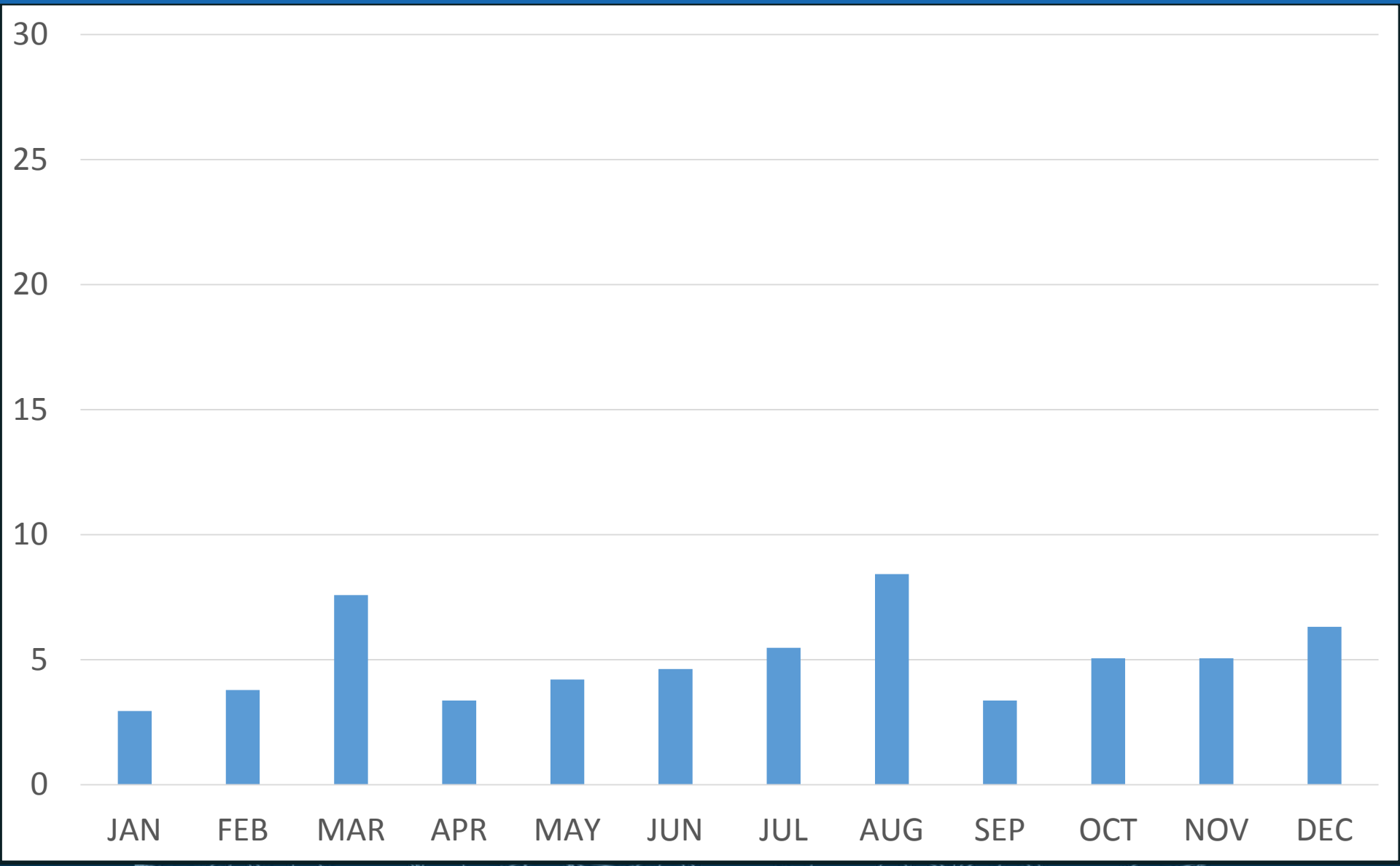
A “validated complaint” is a complaint received by the ICANN compliance department that has been acted on. In other words, this is not an obviously frivolous complaint.

Whois (In-)Accuracy In Registries

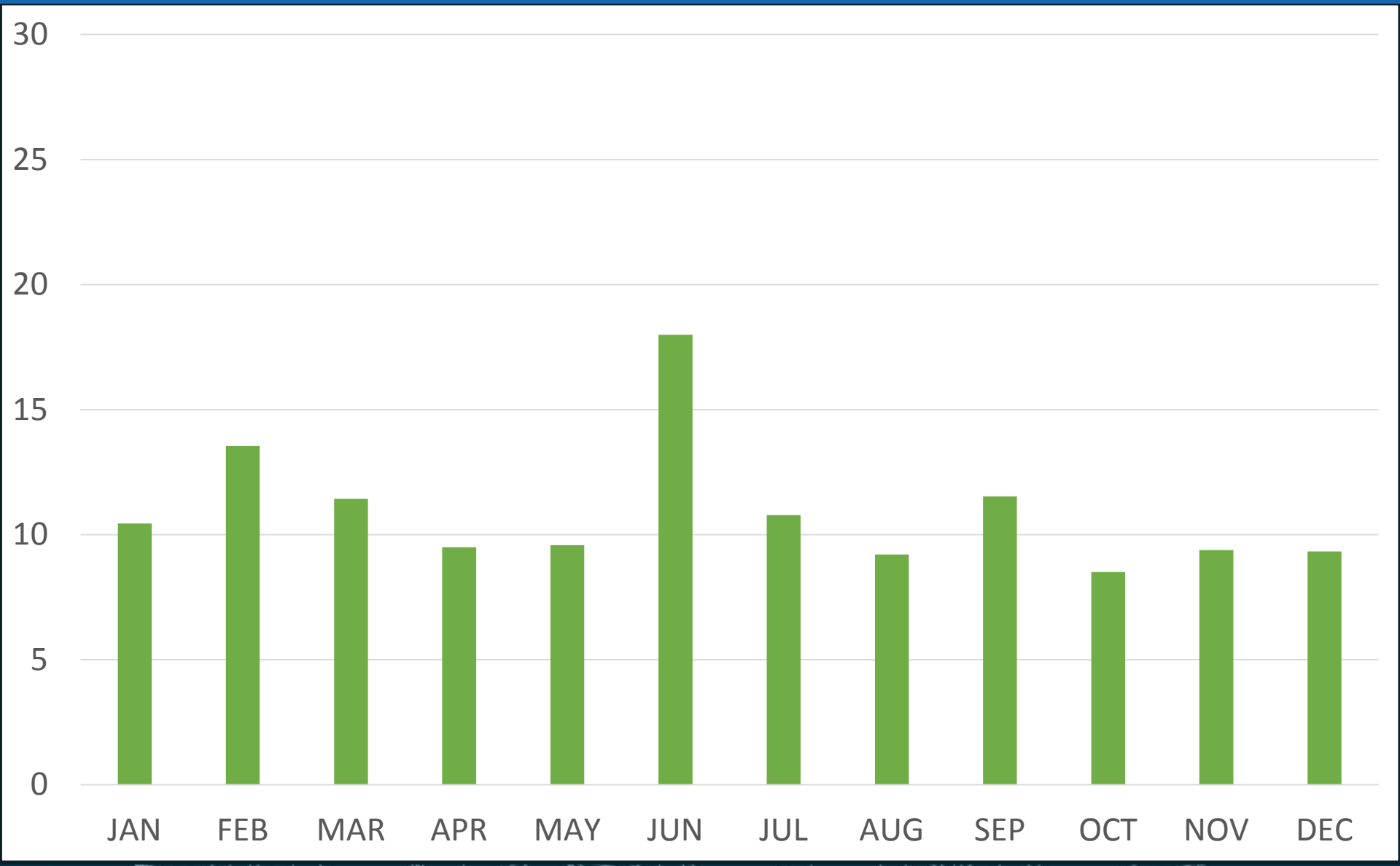
Registriess: Complaints per Million Registrations



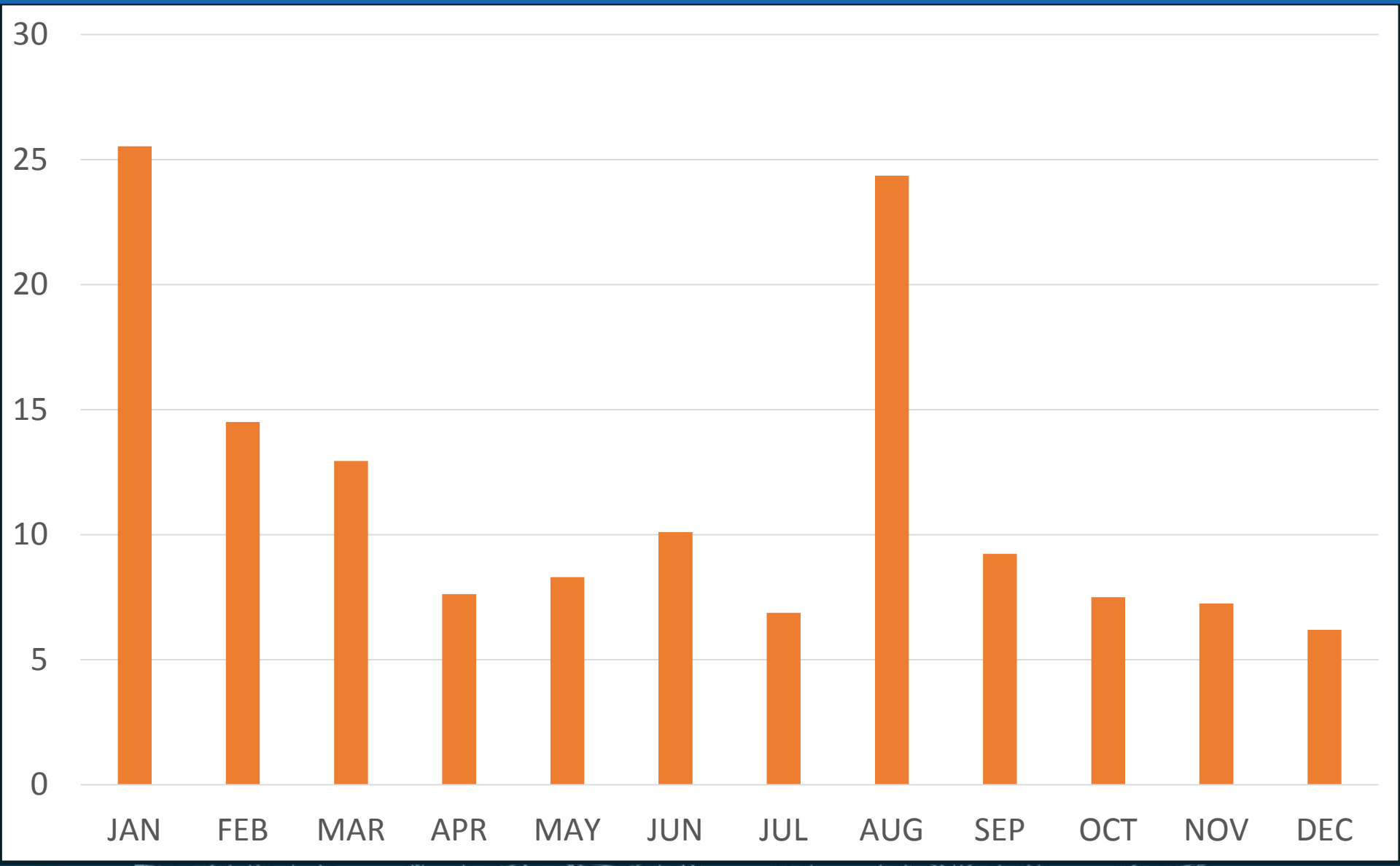
Registry 1: Complaints per Million Registrations



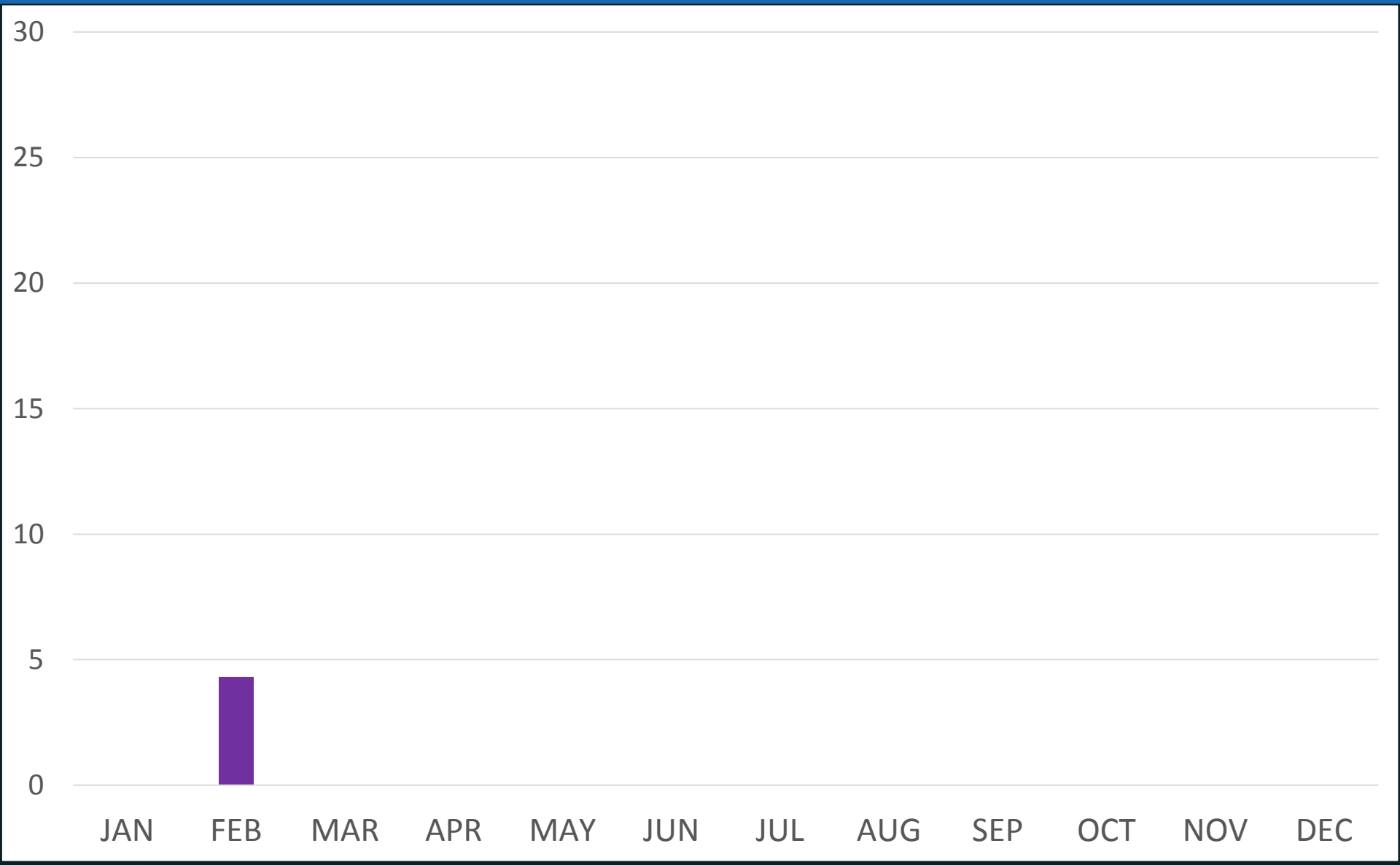
Registry 2: Complaints per Million Registrations



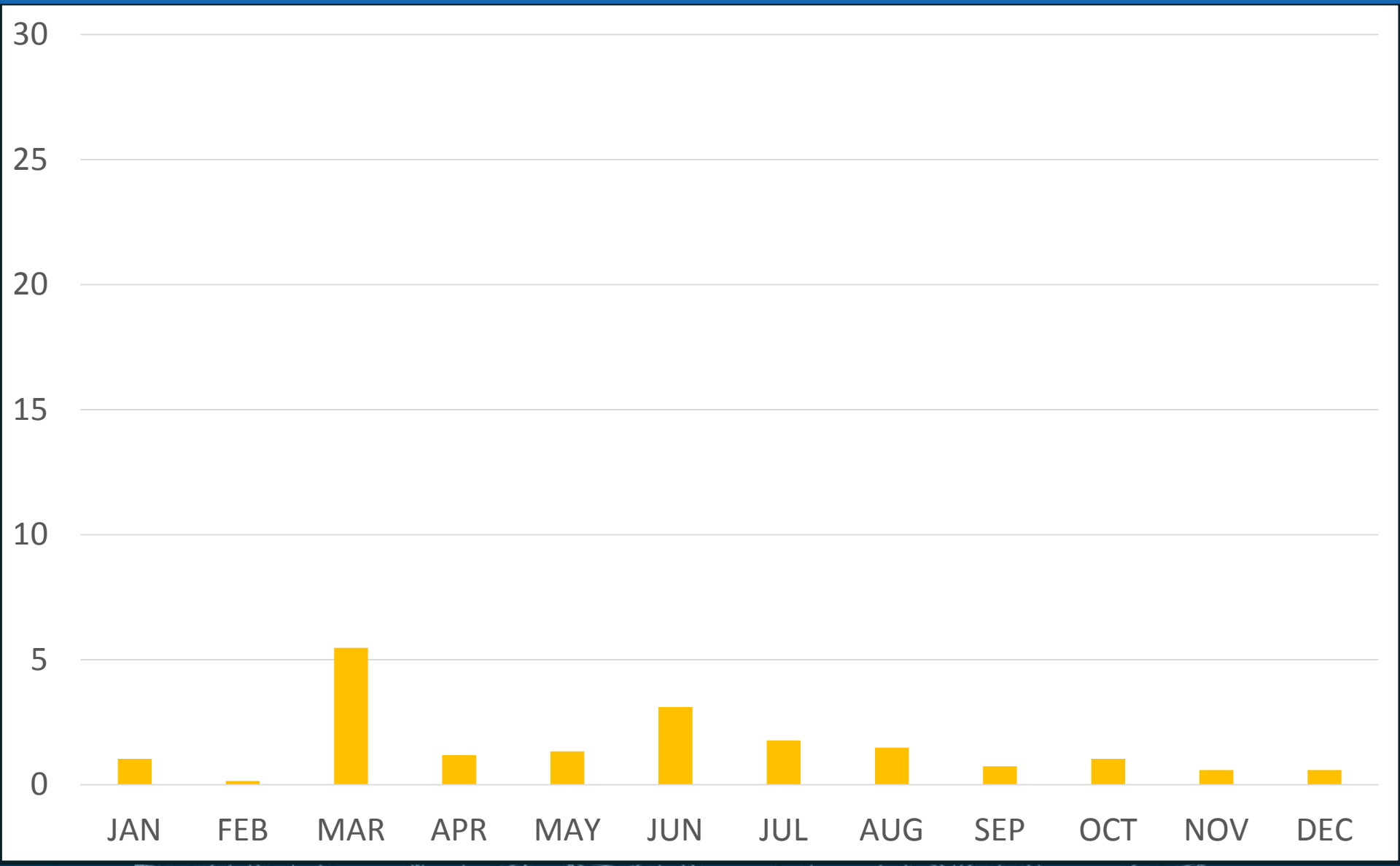
Registry 3: Complaints per Million Registrations



Registry 4: Complaints per Million Registrations

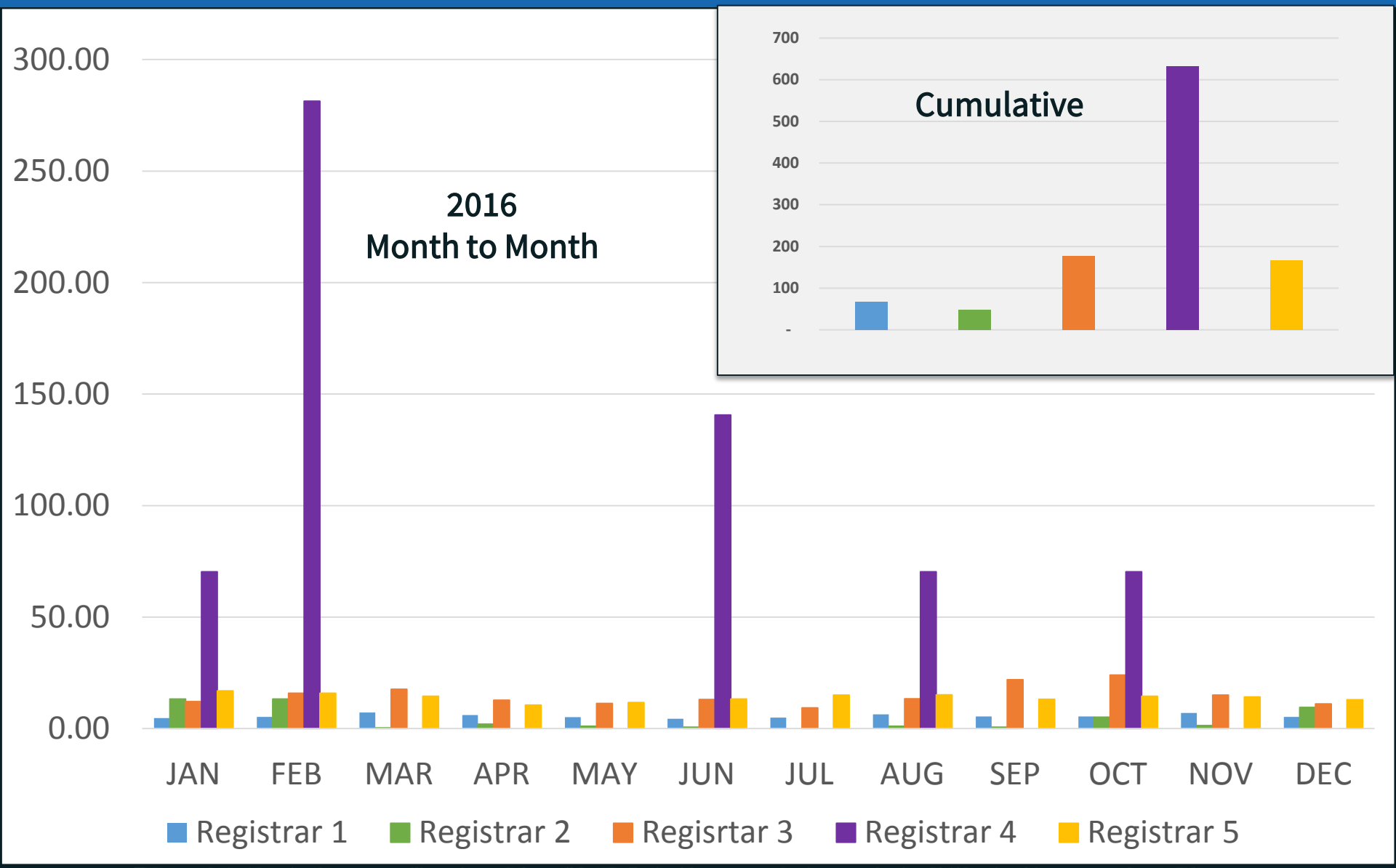


Registry 5: Complaints per Million Registrations

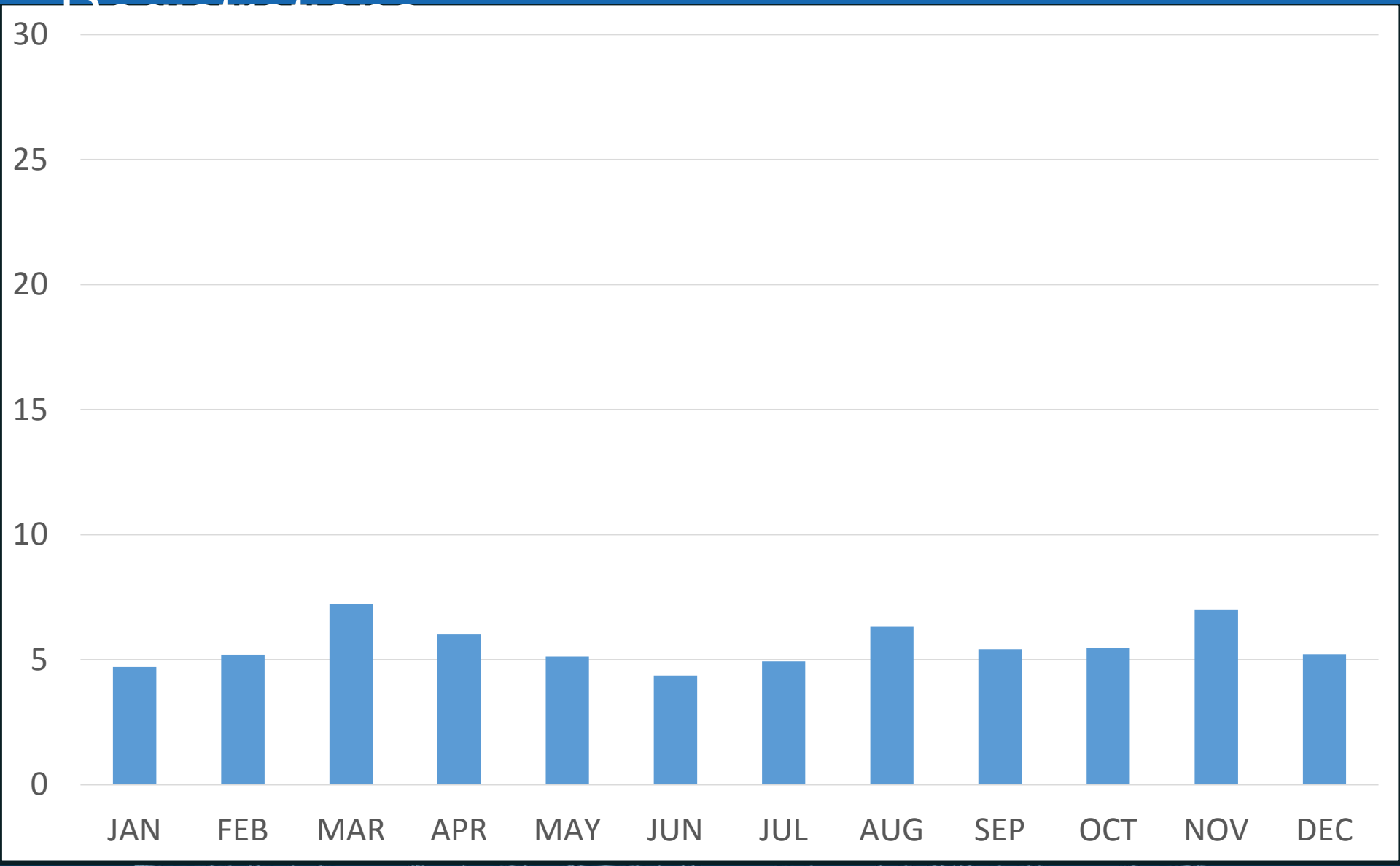


Whois (In-)Accuracy In Registrars

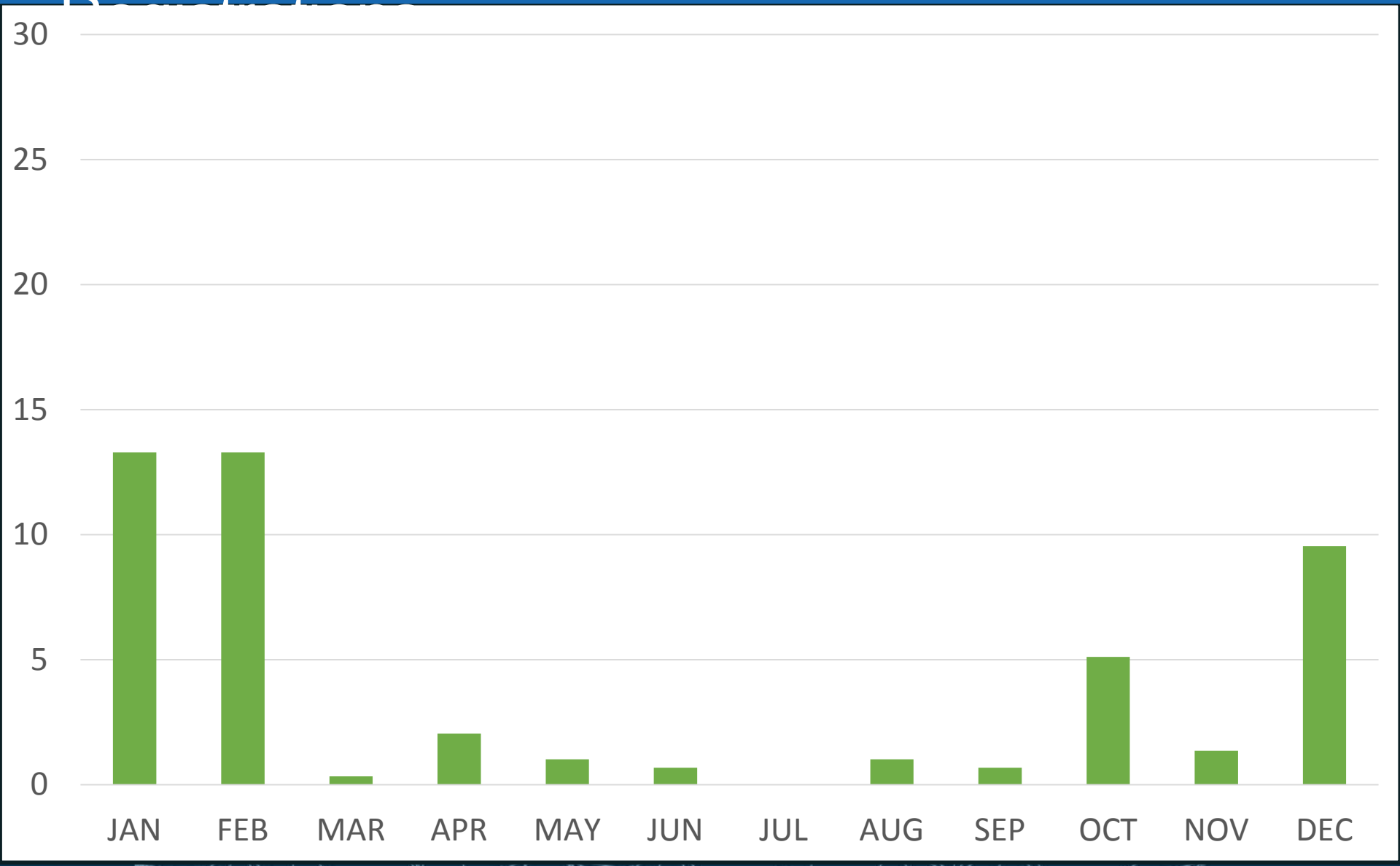
Registrars: Complaints per Million Registrations



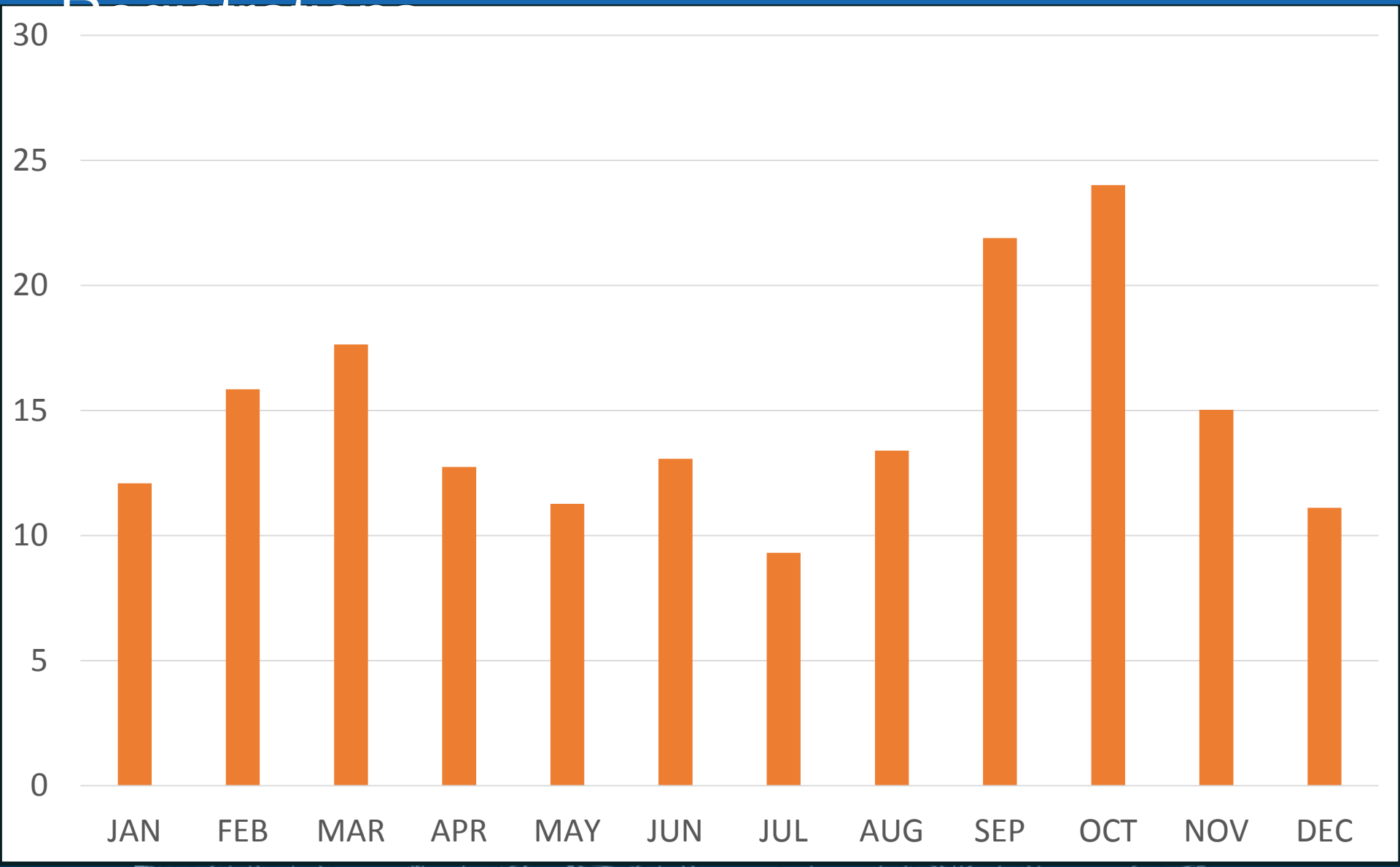
Registrar 1: Complaints per Million Registrations



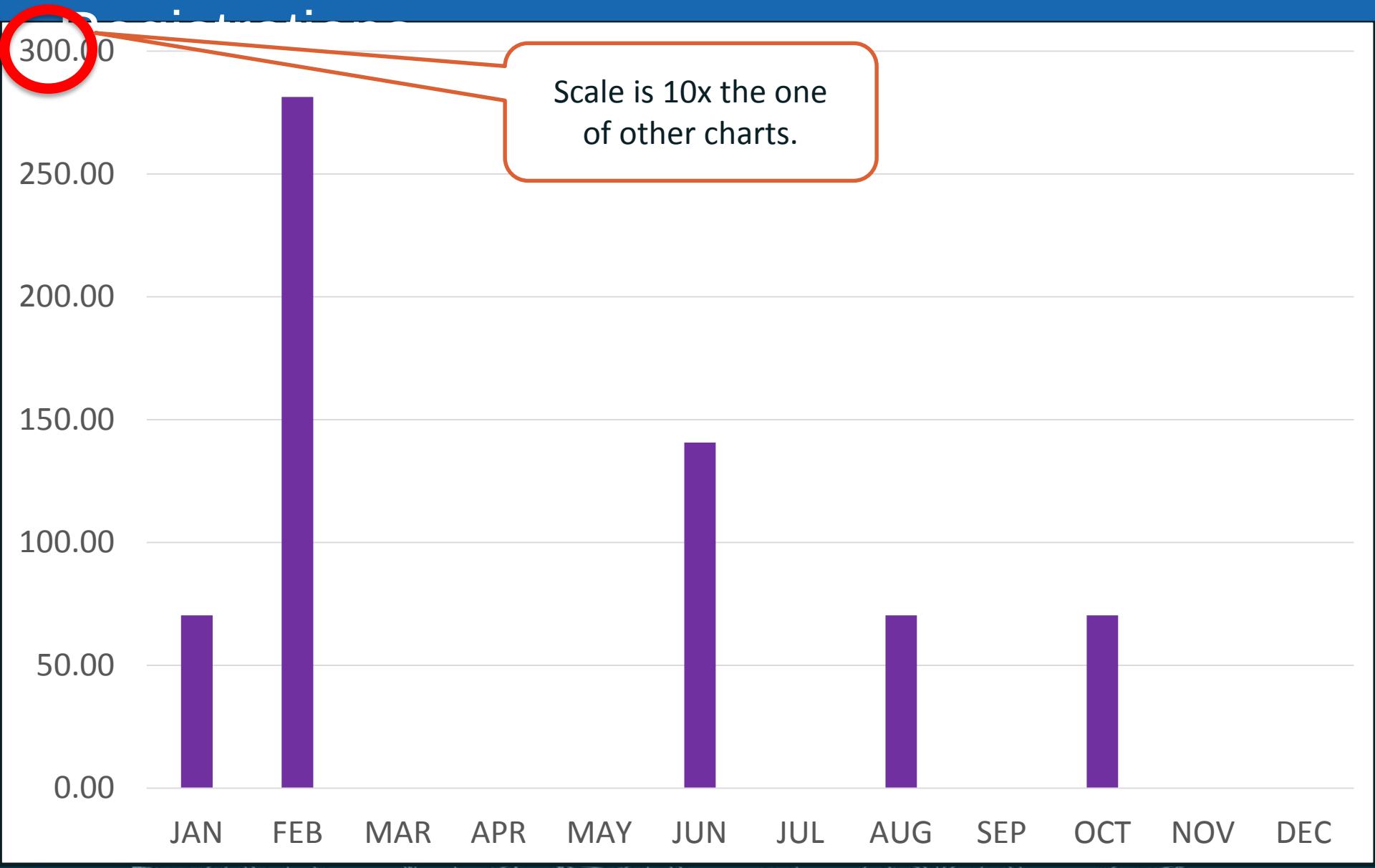
Registrar 2: Complaints per Million Registrations



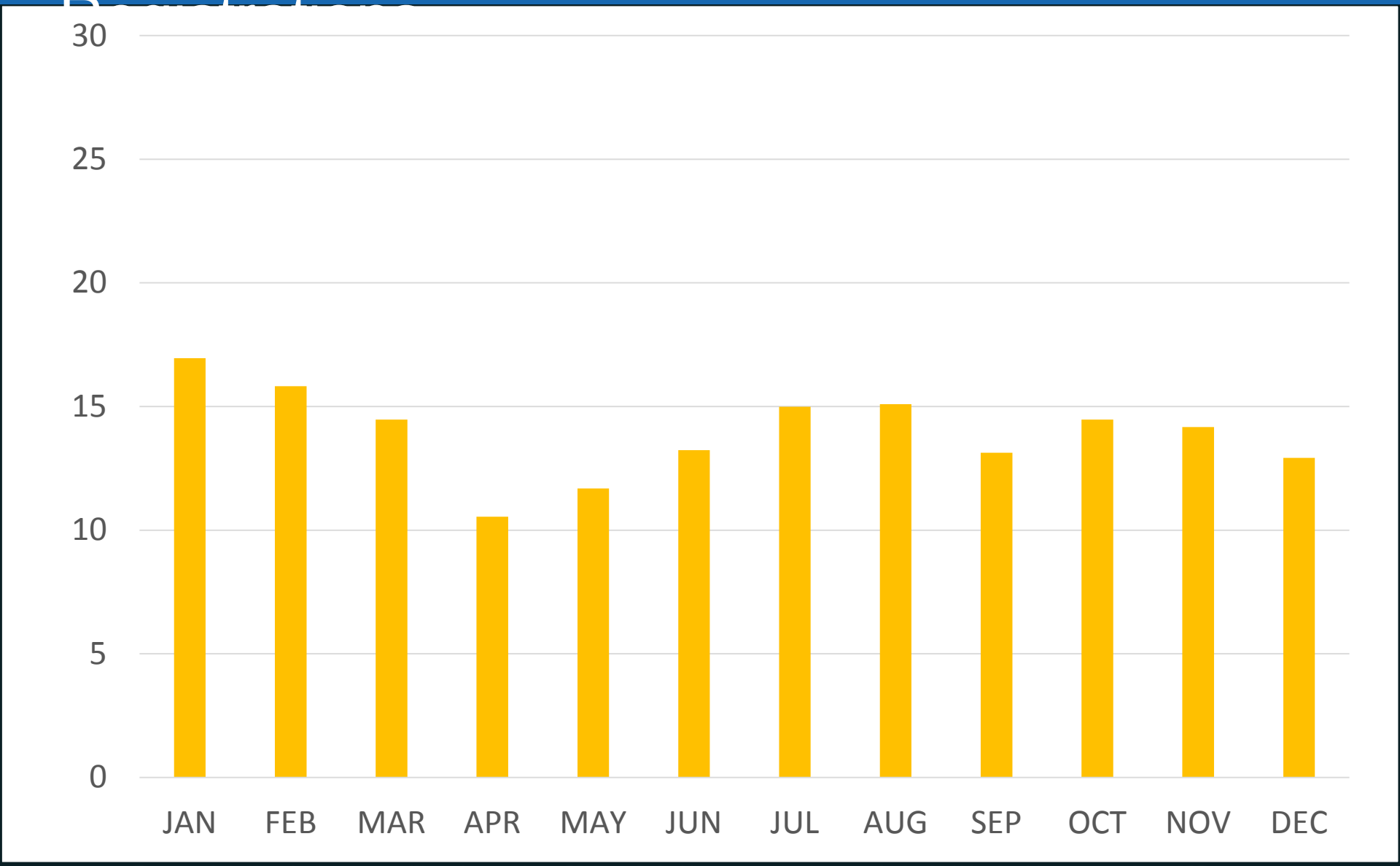
Registrar 3: Complaints per Million Registrations



Registrar 4: Complaints per Million Registrations



Registrar 5: Complaints per Million Registrations



Observations

- The number of complaints received per registrar or registry is relatively small. Typically less than 1 per day or a couple per week on a monthly average.
- There are some exceptions, where we see peaks up to 10 per day on a monthly average.
- We tend to see more differences among the registrars than among the registries.
- This is only a sample of 5 Registries and 5 Registrars.

Domain Name Abuse

ITHI Cooperation with SSR

We worked in conjunction with the DNS Abuse Reporting Tool (DART) to develop a set of domain name abuse candidate metrics M2.

DART is based on a number of industry accepted feeds.

Data is available since November 2016. In this prototype, we use only one data point for the same registries as previous study.

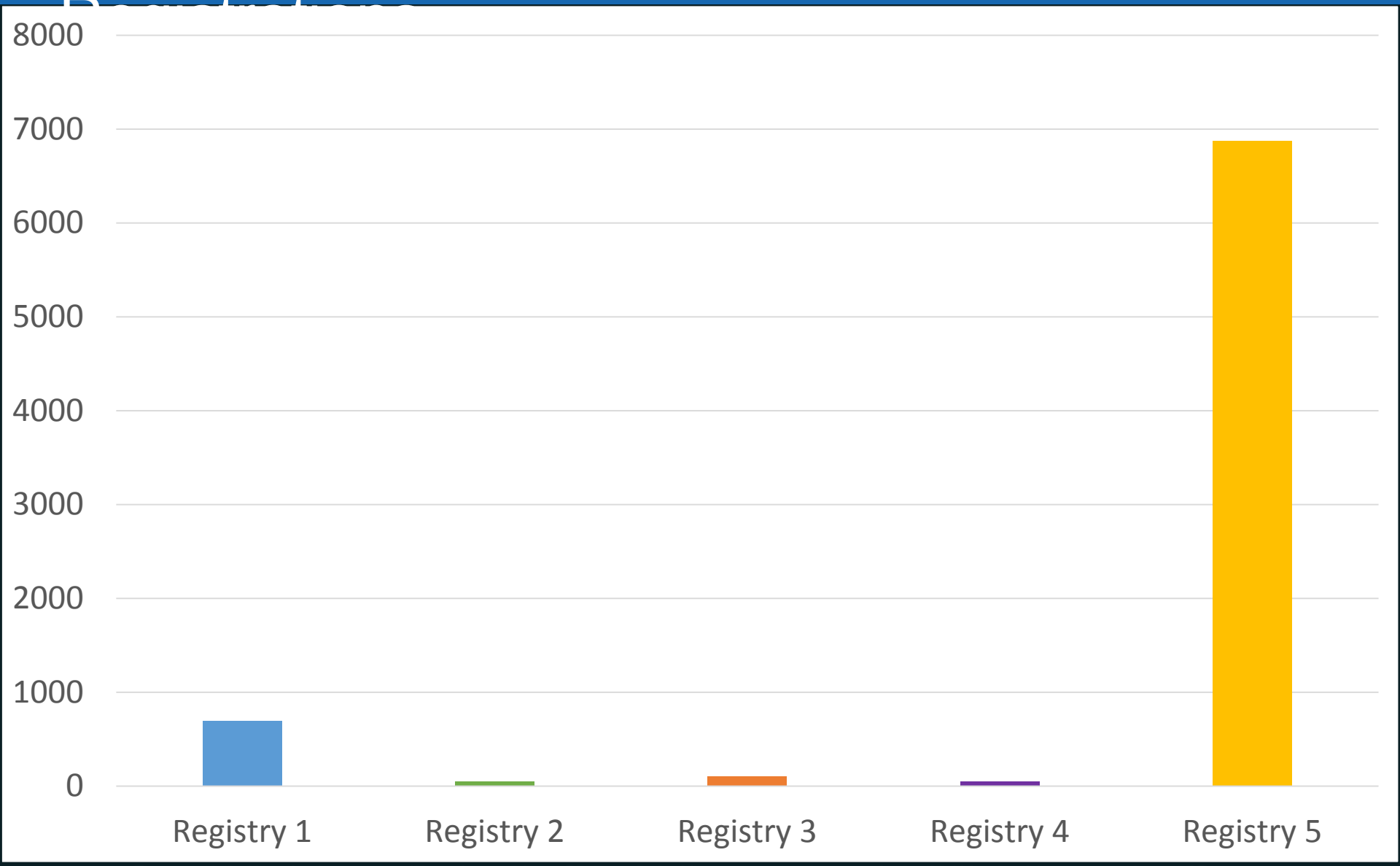
Because this is only a limited sample and the methodology is still under development, we have anonymized the data to avoid singling out anybody.

Candidate Metric Related to Abuse

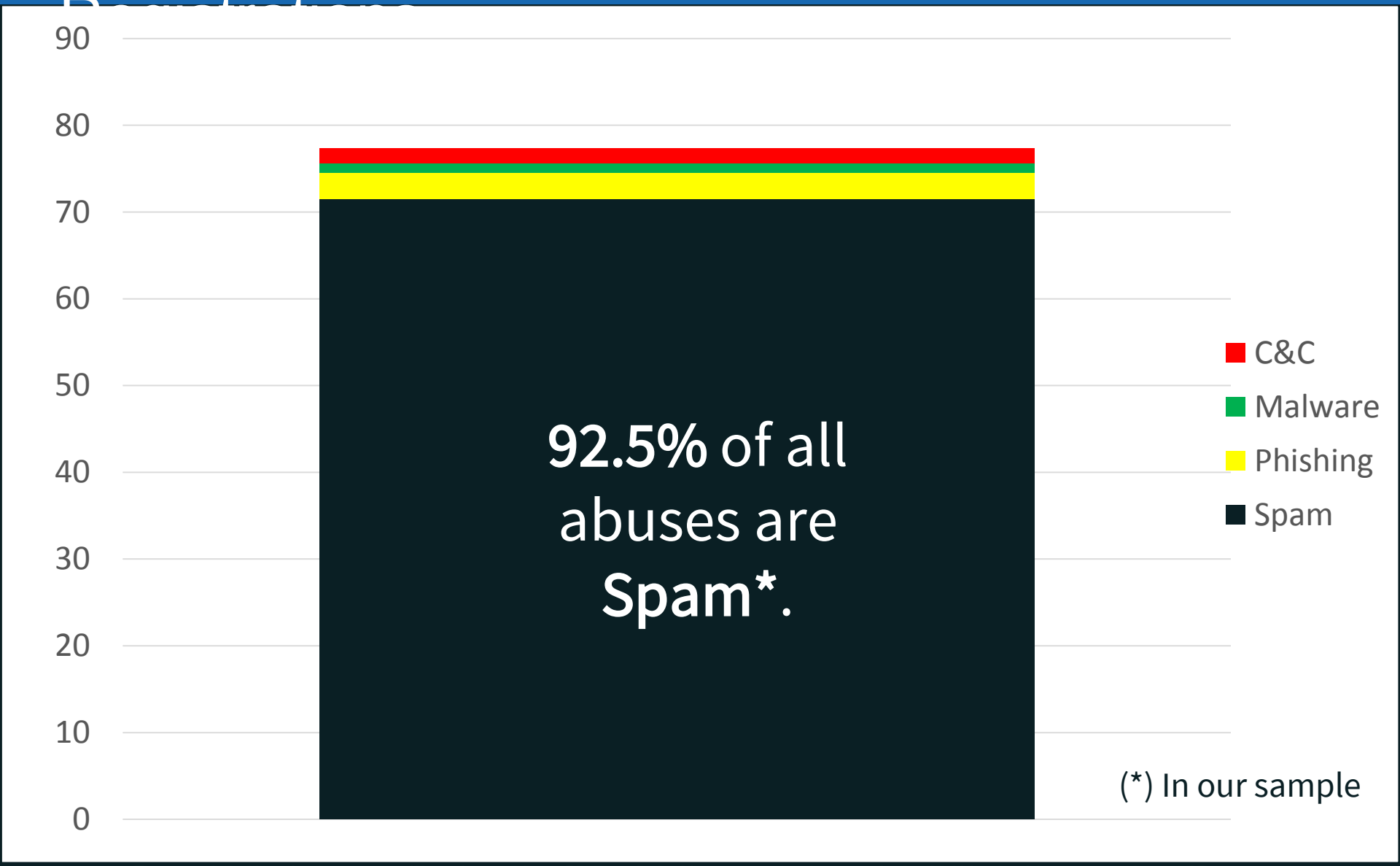
M2	Number of abuses in the feeds per 10,000
M2 is then broken down in 4 metrics	
M2.1	Spams
M2.2	Phishes
M2.3	Malwares
M2.4	Command & Control

Domain Name Abuse In Registries

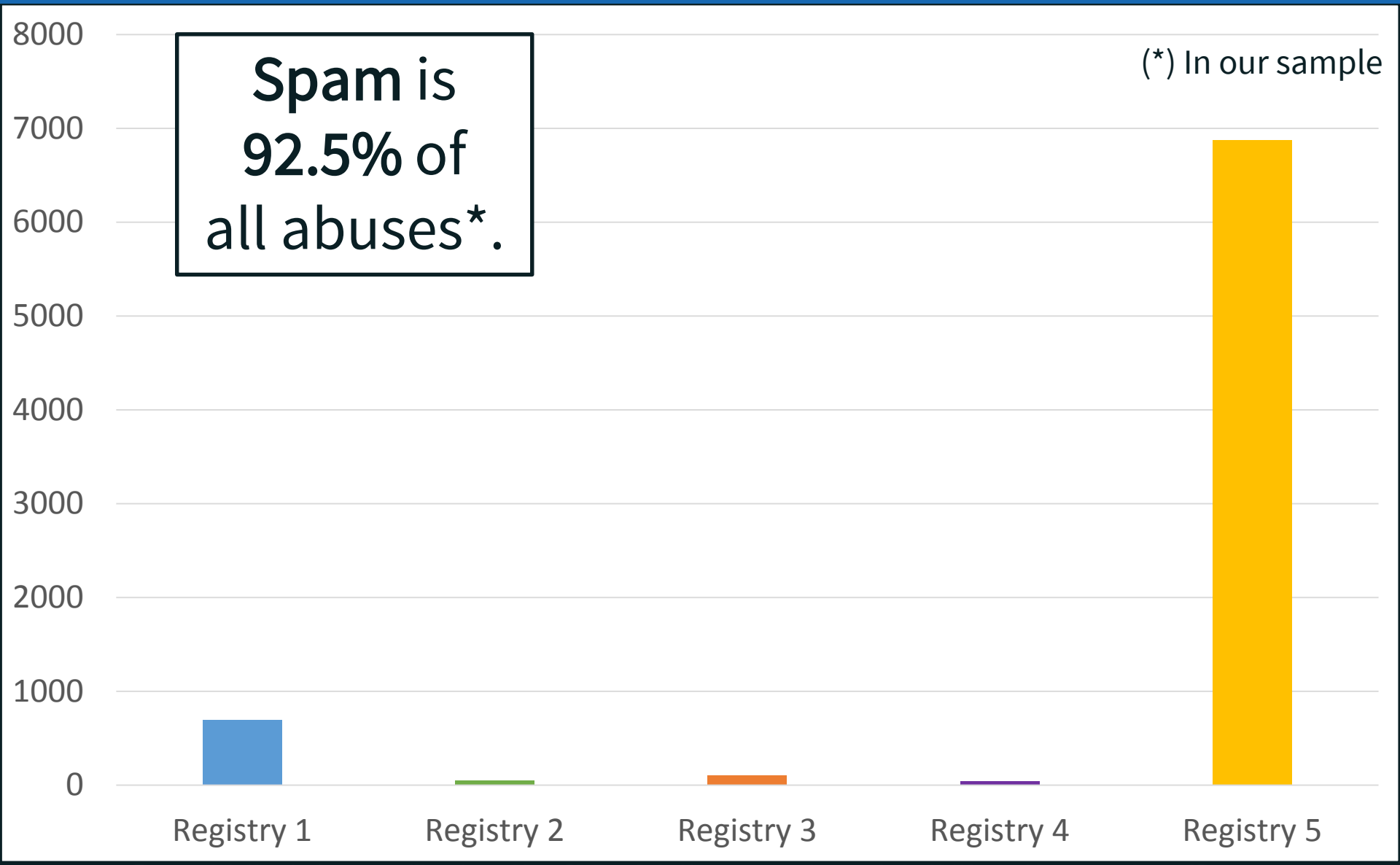
Total Abuses on 2017-05-04 per 10k Registrations



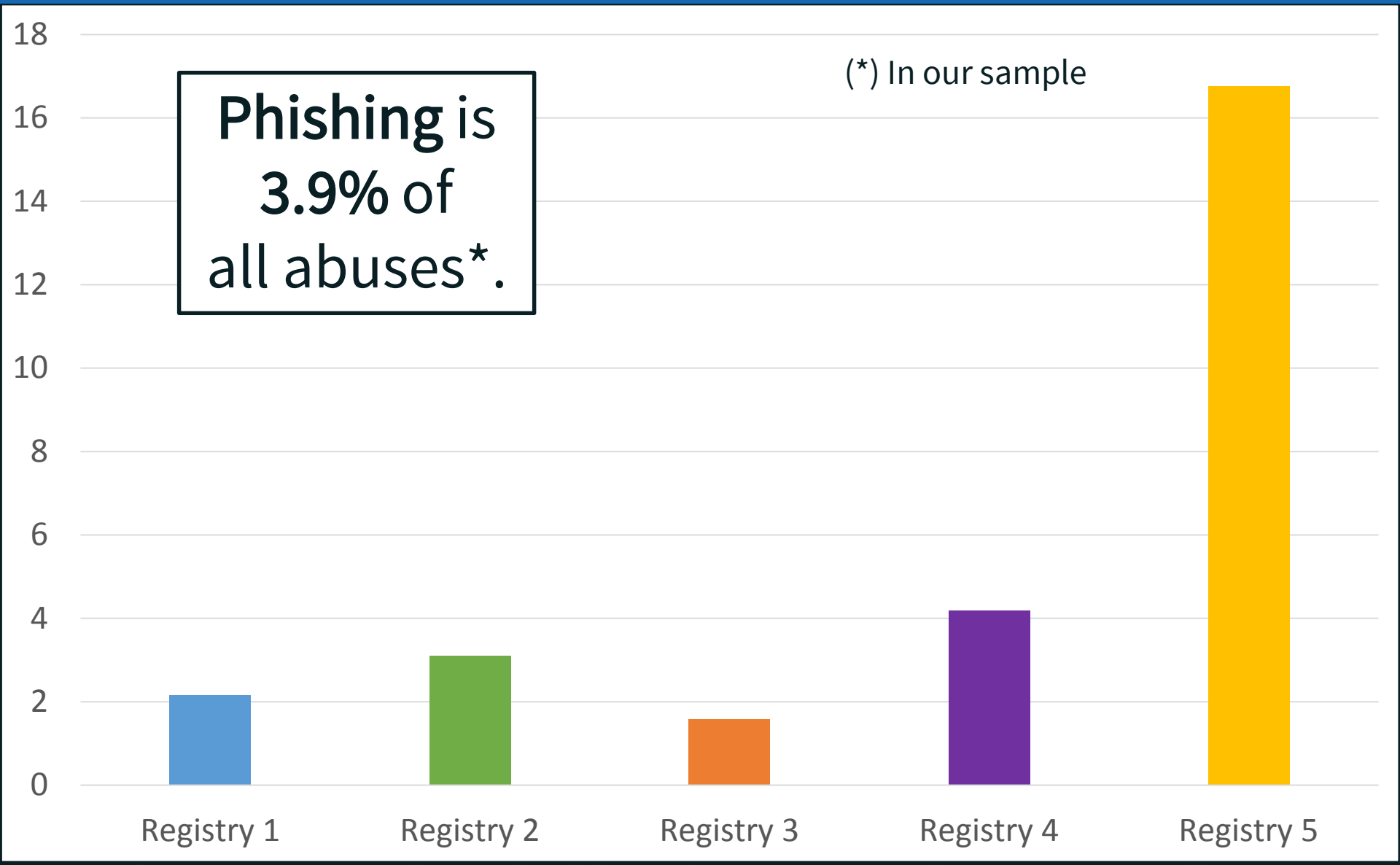
Abuses Across the 5 Registries per 10k Registrations



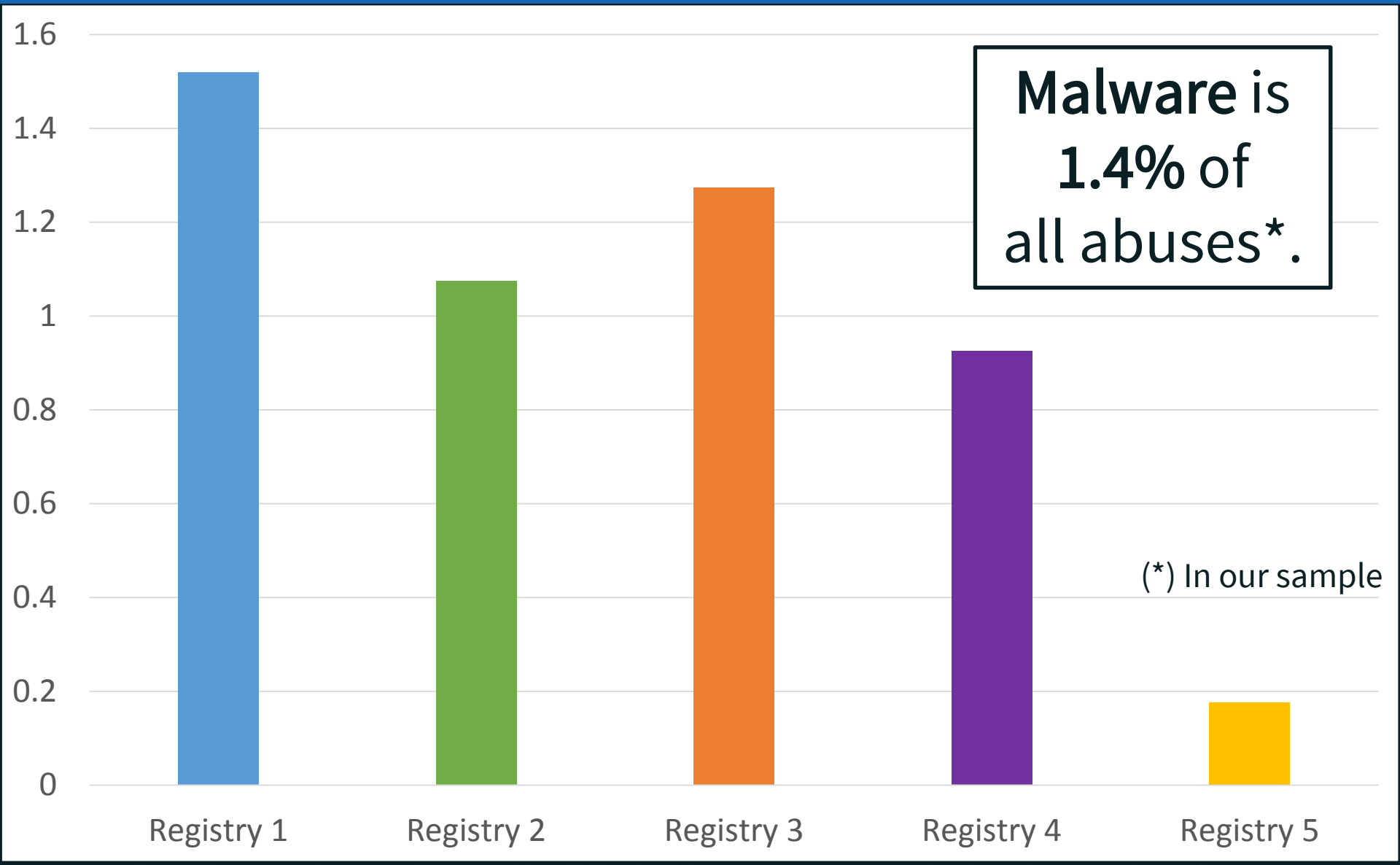
Spam on 2017-05-04 per 10k Registrations



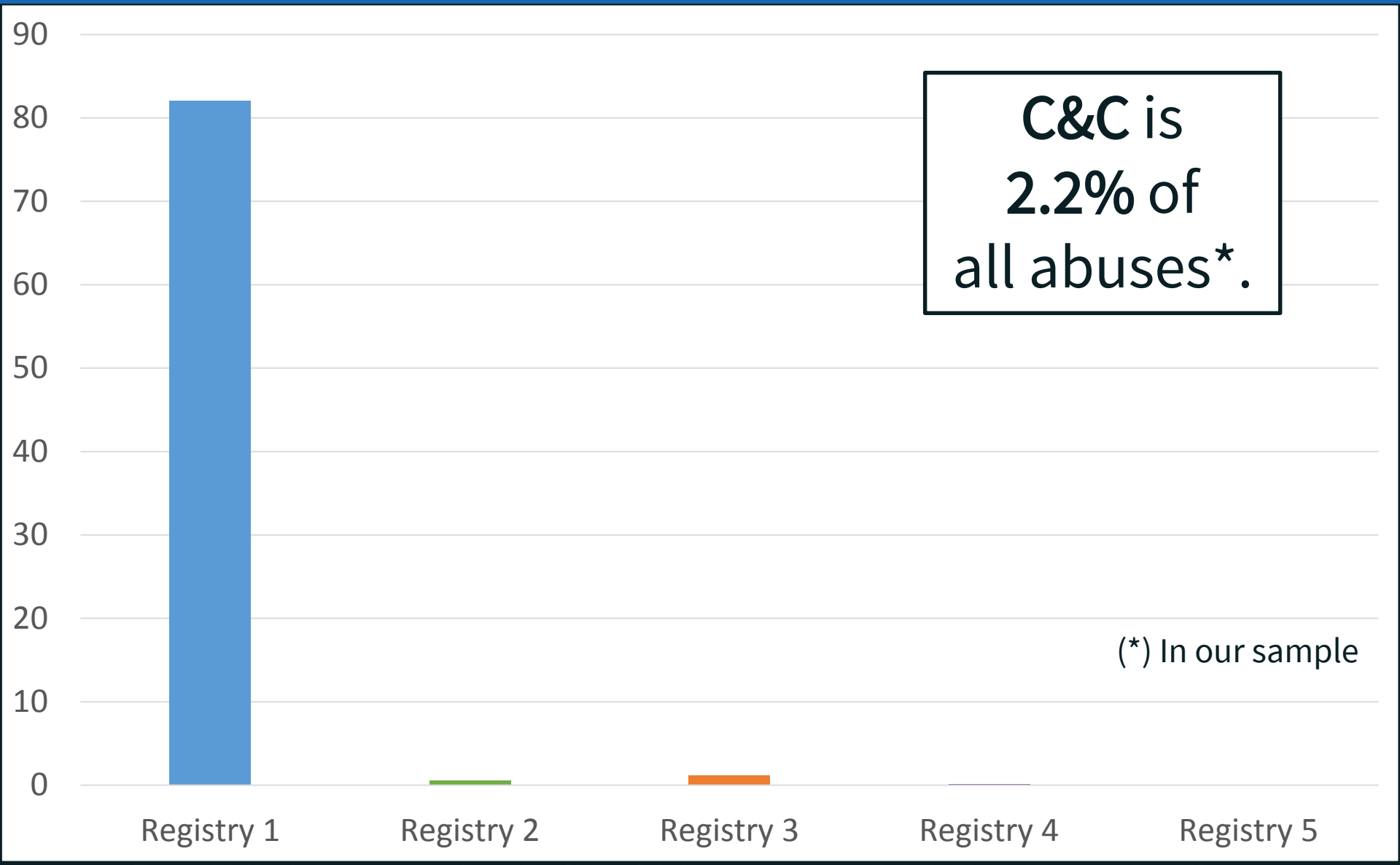
Phishing on 2017-05-04 per 10k Registrations



Malware on 2017-05-04 per 10k Registrations



C&C on 2017-05-04 per 10k Registrations



Observations

- There is much more data to be used in the cumulated abused feeds than in whois inaccuracy complaints.

- Spam, Phishing, Malware, Command & Control are not affecting all registries and registrars equally.
  Spam is by far the largest problem:
  - **Up to 95% in our sample data**

- **Significantly different abuse profiles are emerging among registries and registrars**.
  - This study is only covering 5 registries and 5 registrars. We can extend it to cover them all.