However, we don't know the optimal dosage to give to patients.

Subheading goes here
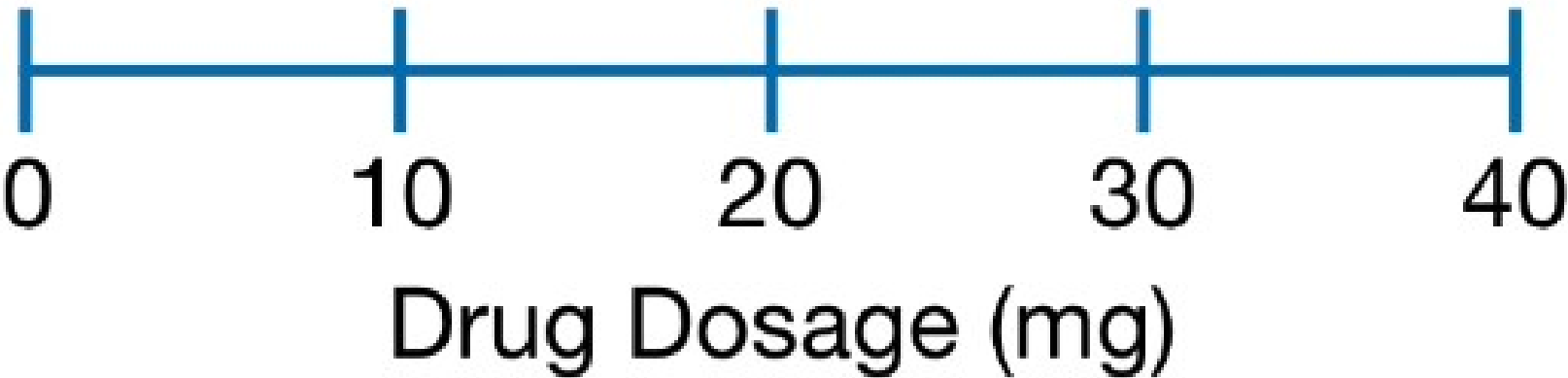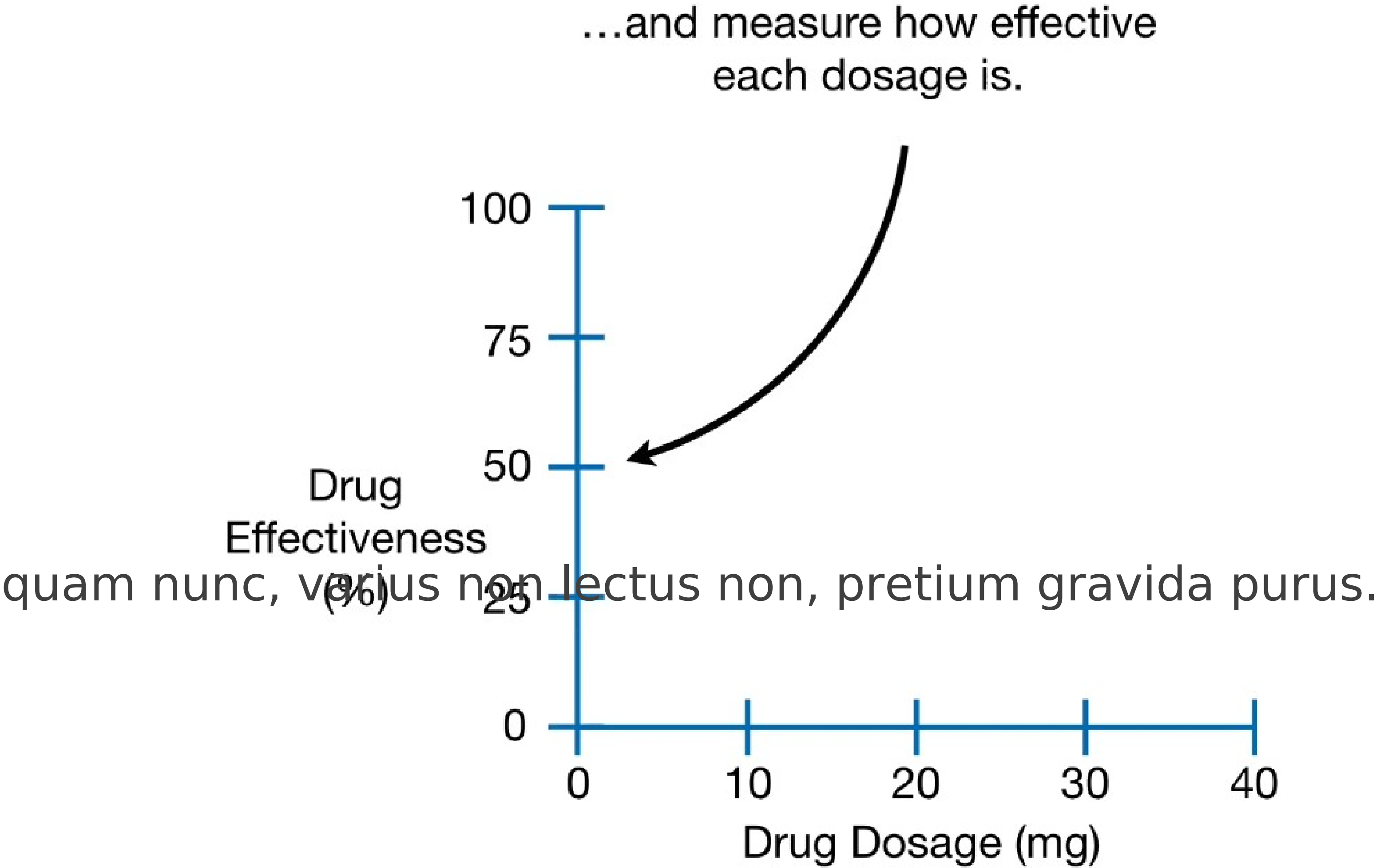

VS.

Developer Student Clubs

Google Developers

So we do a clinical trial with
different dosages…

d

cu

Sed

quam nunc, varius non lectus non, pretium gravida purus.

0   10   20   30   40
Drug Dosage (mg)

...and measure how effective
each dosage is.

id

ʿcu

. Sed

quam nunc, varius non lectus non, pretium gravida purus.

100

75

50

Drug
Effectiveness
(%)   25

0

0   10   20   30   40

Drug Dosage (mg)

If the data looked like this…

Drug
Effectiveness
(%)

100

75

50

25

0

0   10   20   30   40

Drug Dosage (mg)

quam nunc, varius non lectus non, pretium gravida purus.

...and, in general, the higher the dose,

Drug Effectiveness (%)

100
75
50
25
0

0   10   20   30   40

Drug Dosage (mg)

e id

arcu

ur. Sed

quam nunc, varius non lectus non, pretium gravida purus.
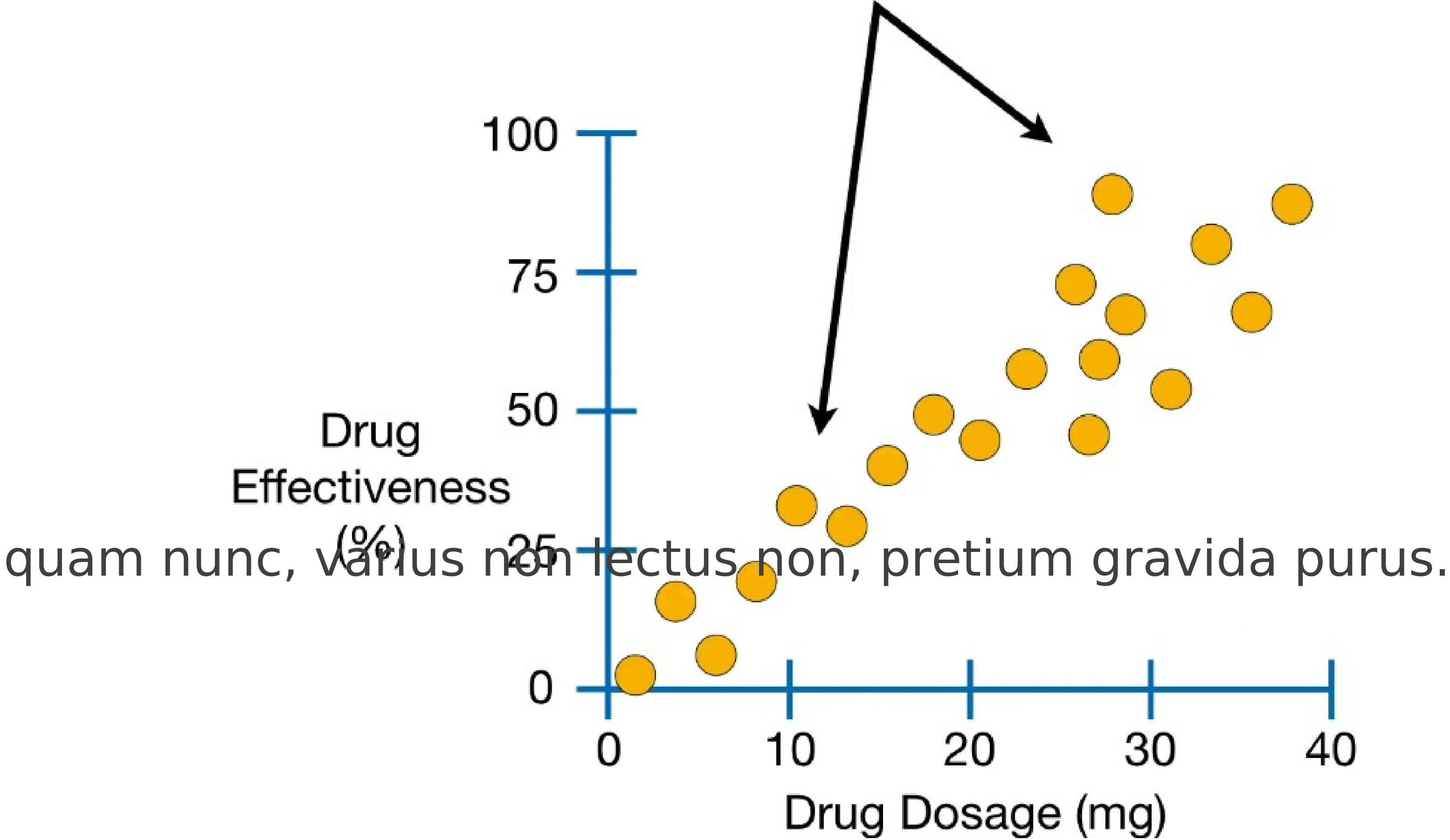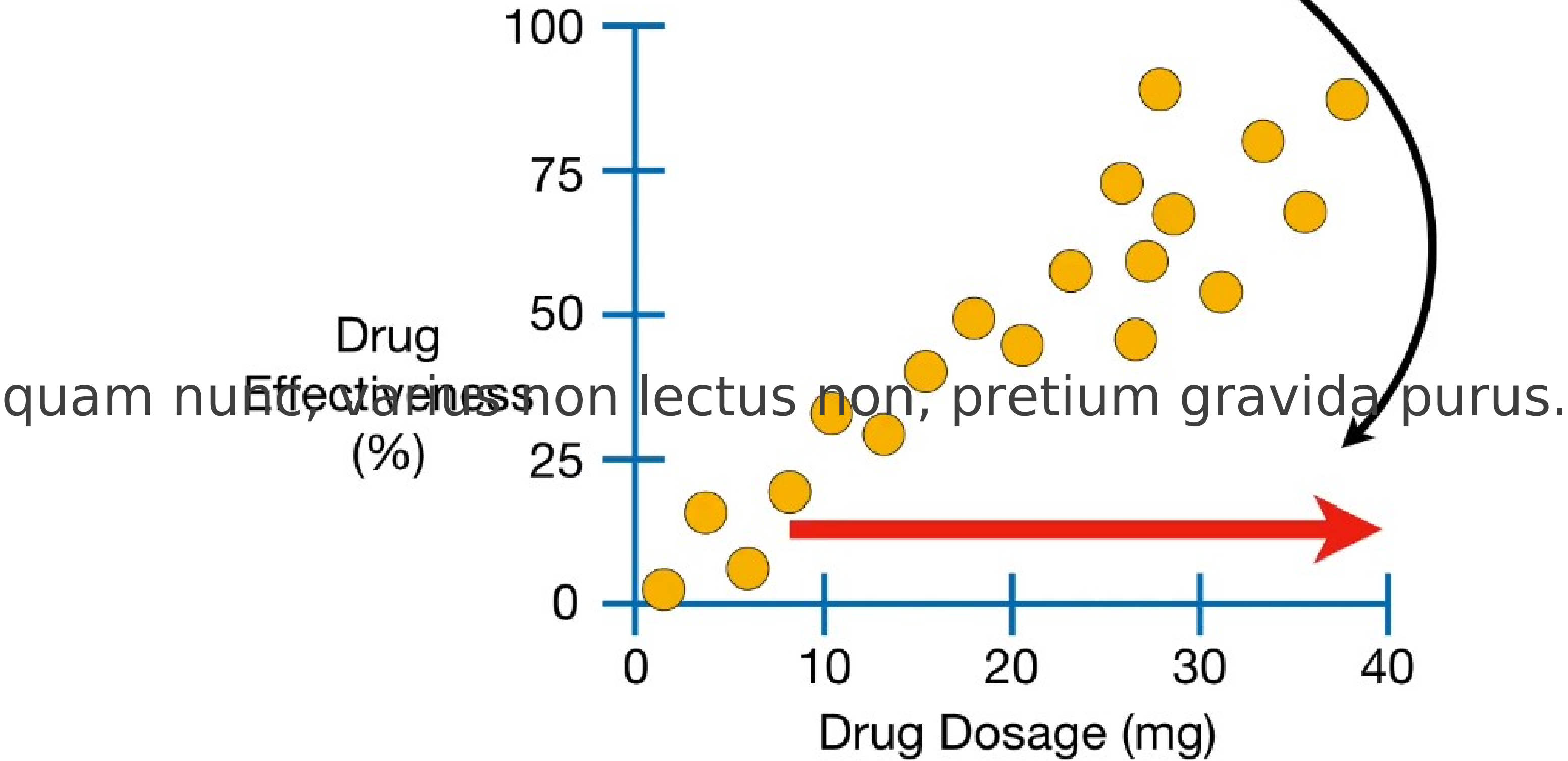
...and, in general, the higher the dose,
the more effective the drug...

Drug
Effectiveness
(%)

Drug Dosage (mg)

e id

t

arcu

ur. Sed

quam nunc, varius non lectus non, pretium gravida purus.
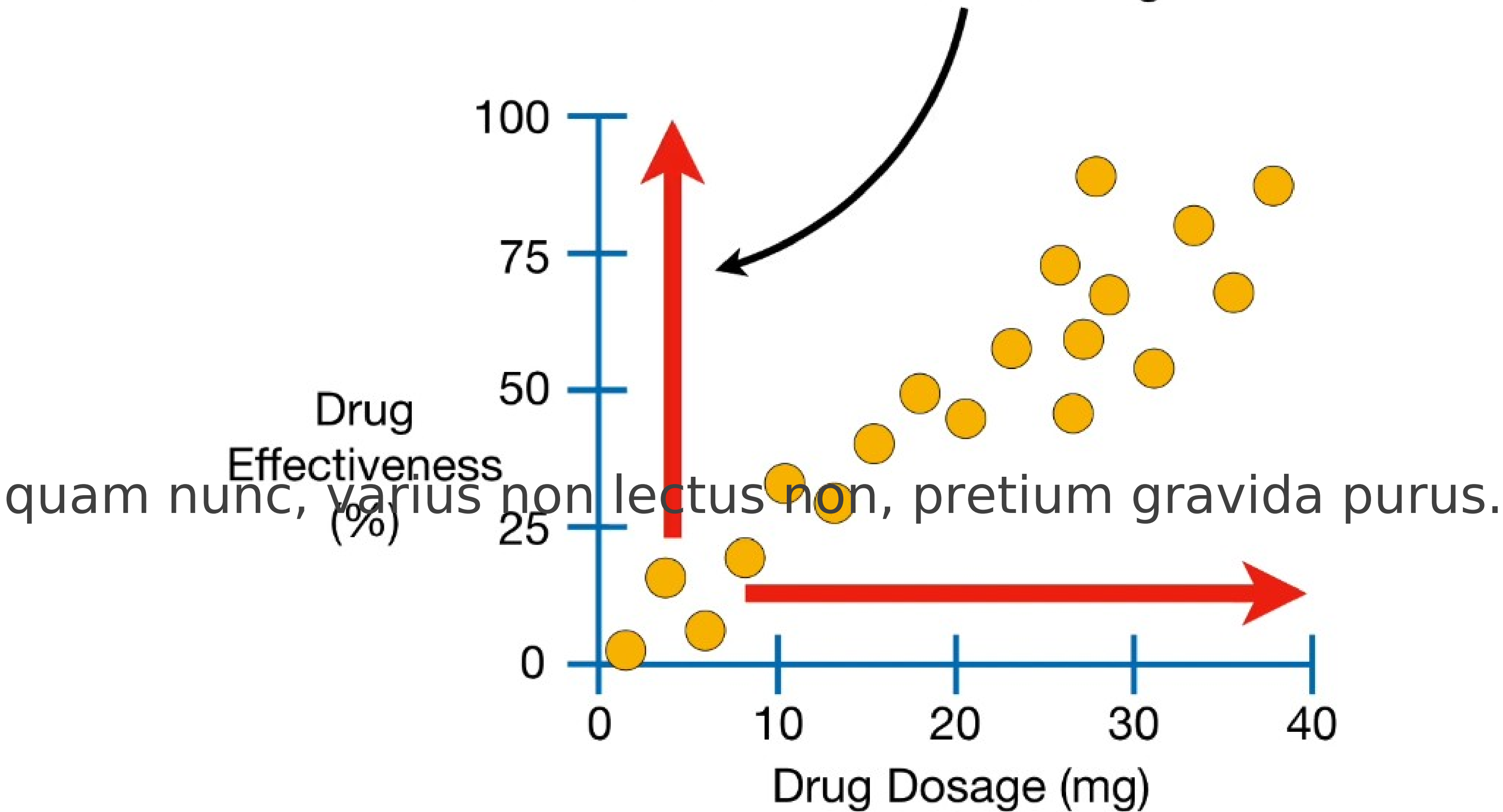
...we could use the line to predict that a
27 mg Dose should be **62% Effective.**

quam nunc, varius non lectus non, pretium gravida purus.

u

Sed
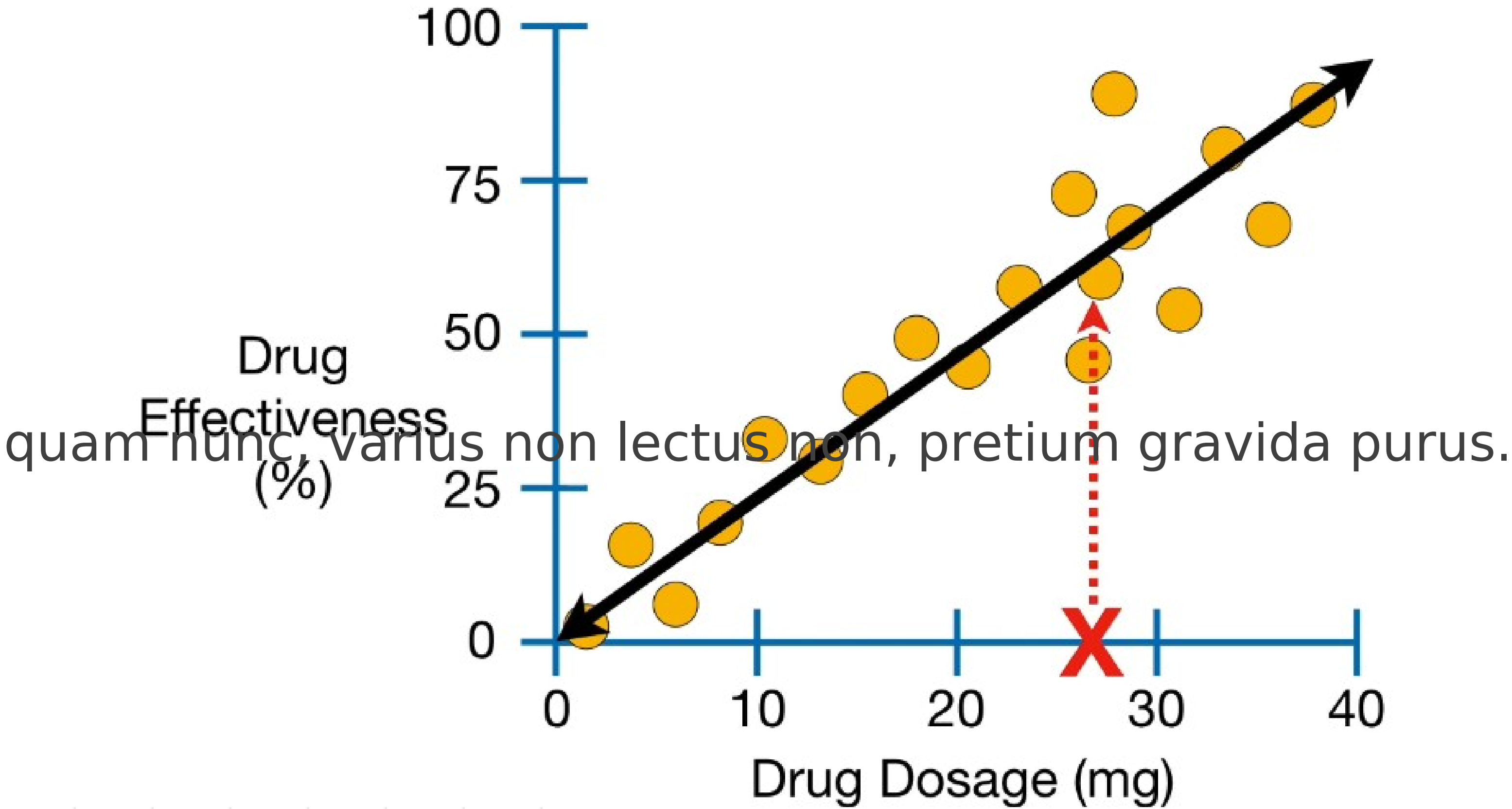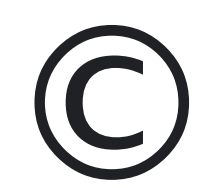
However, what if the data
looked like this?

100

75

id

Drug
Effectiveness

50

rcu

. Sed

quam nunc, varius non lectus non, pretium gravida purus.
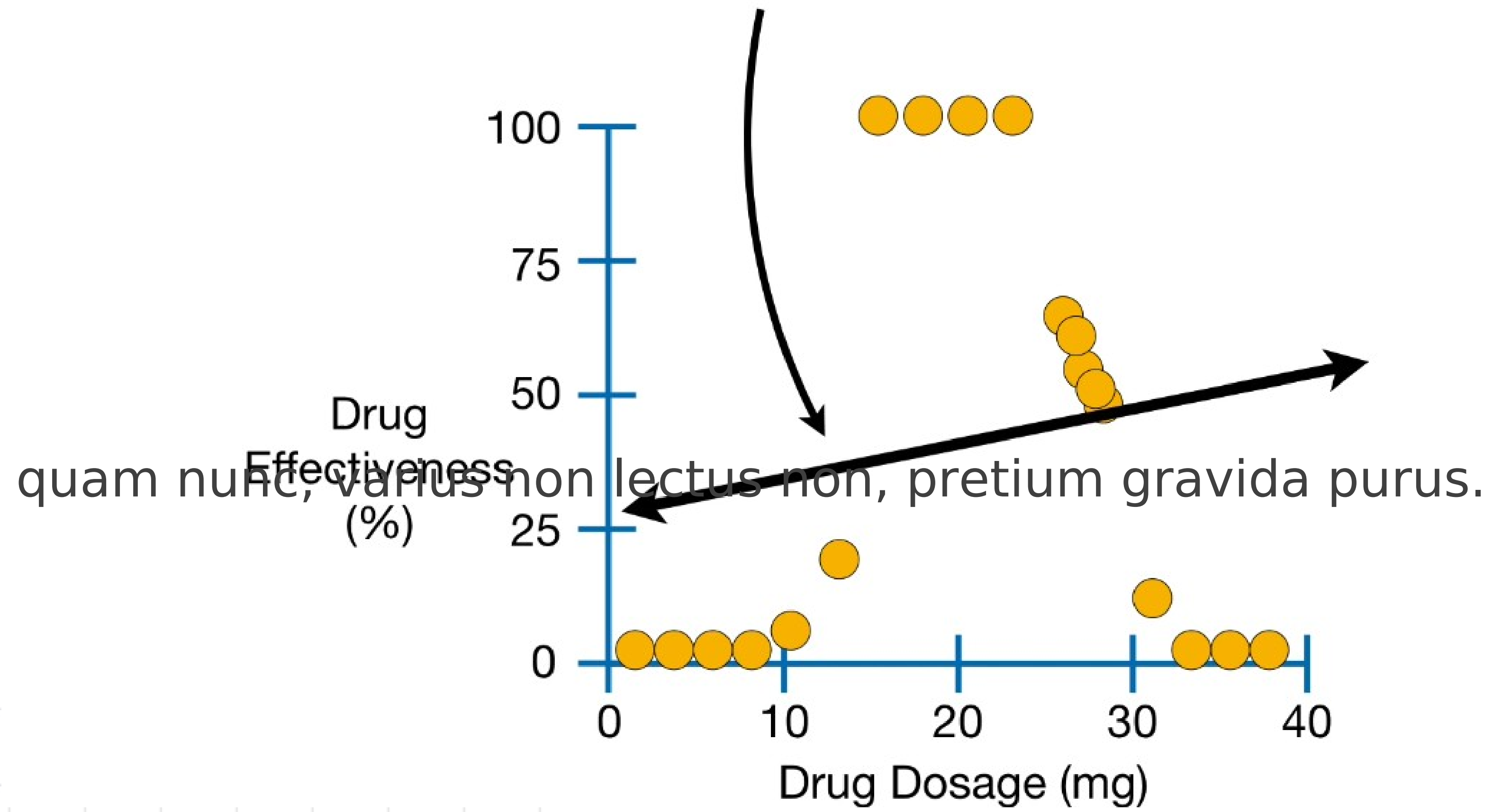
25

0

0   10   20   30   40
Drug Dosage (mg)
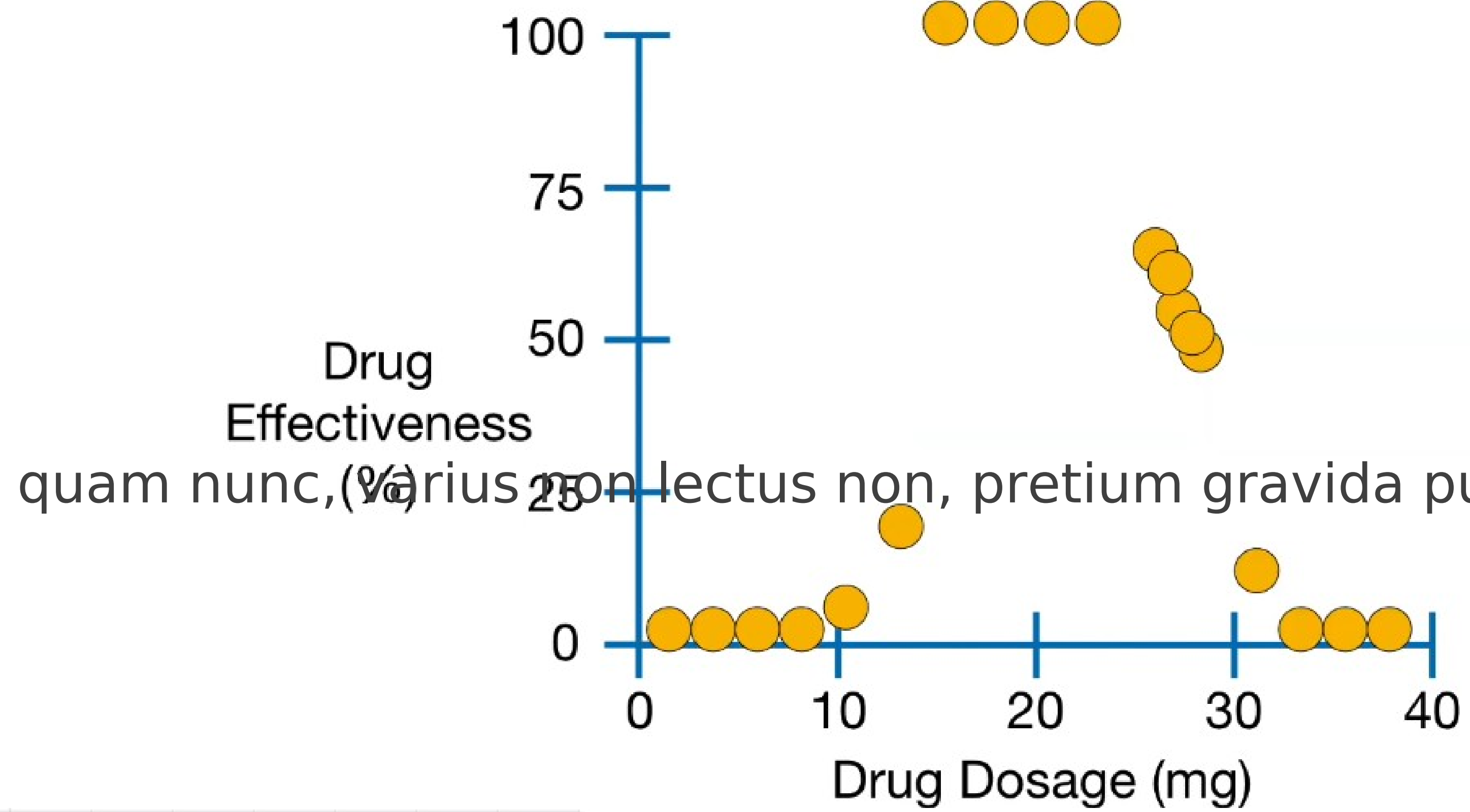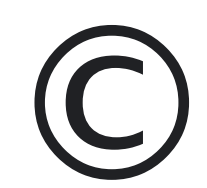
In this case, fitting a straight line
to the data will not be very useful.

re id

t

arcu

tur. Sed

quam nunc, varius non lectus non, pretium gravida purus.

Drug
Effectiveness
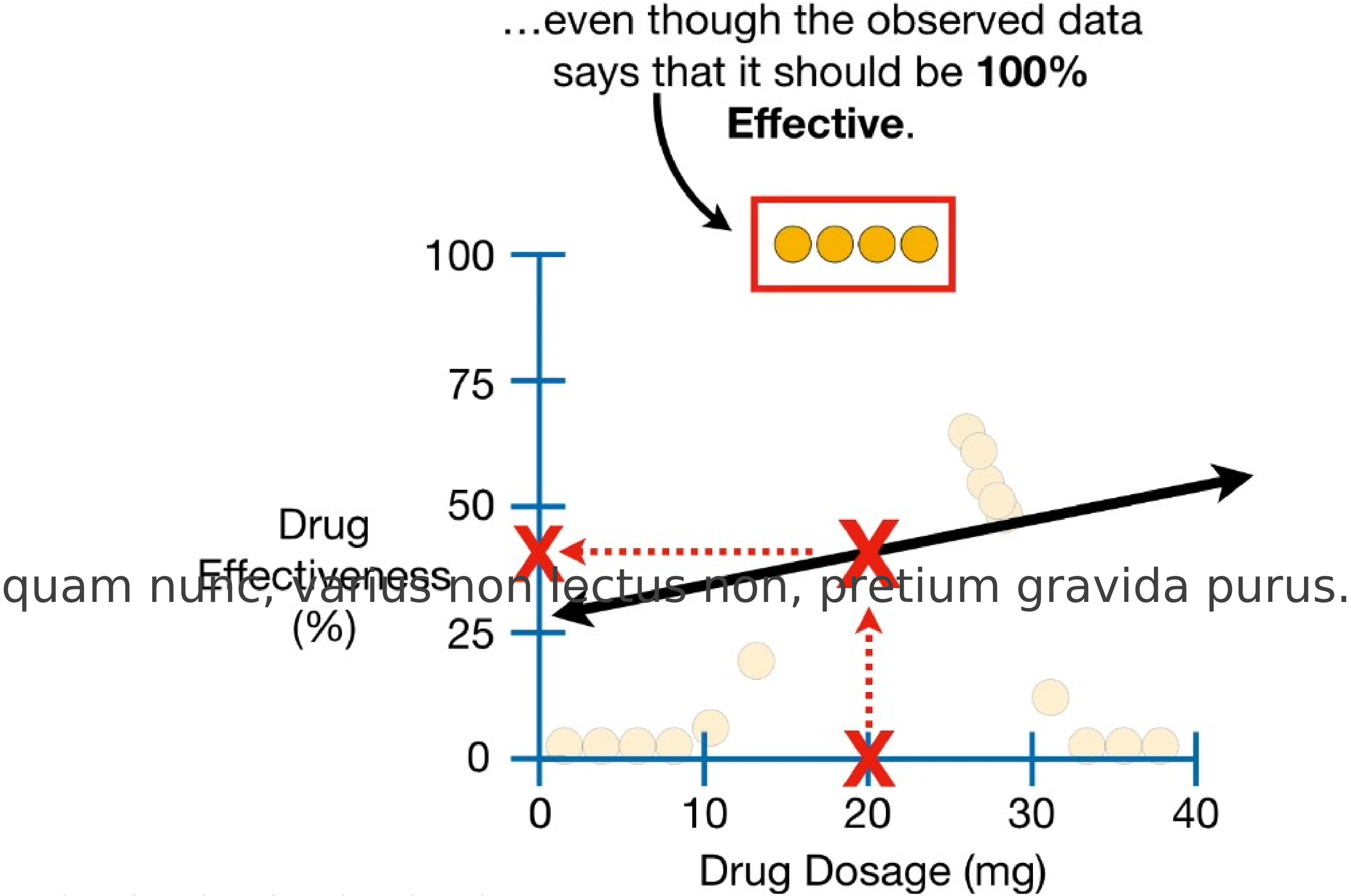(%)

100

75

50

25

0

0   10   20   30   40

Drug Dosage (mg)

...then we would predict that a **20 mg**
Dose should be **45% Effective**...

quam nunc, varius non lectus non, pretium gravida purus.

…even though the observed data
says that it should be **100%**
Effective.

quam nunc, varius non lectus non, pretium gravida purus.

id
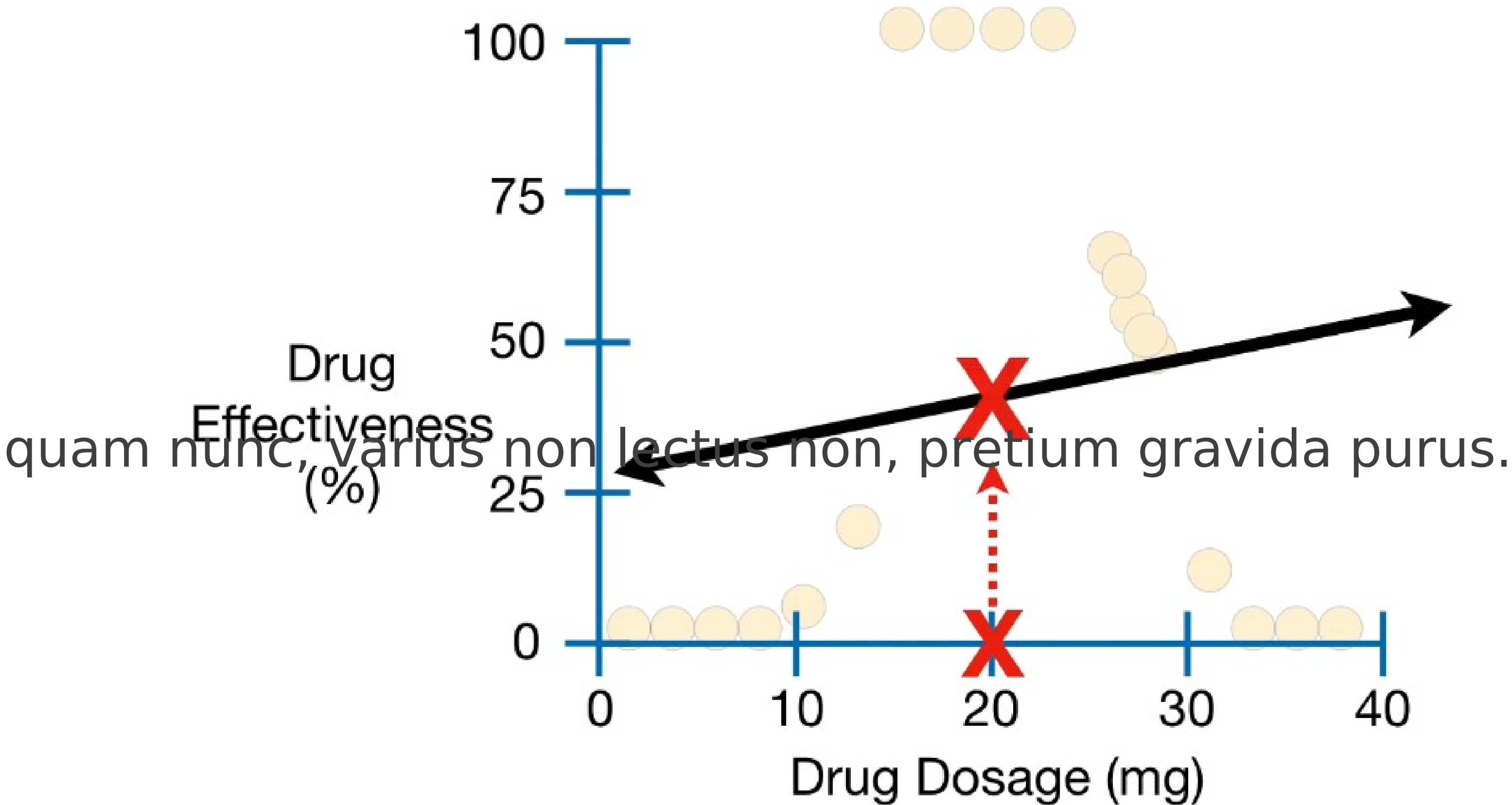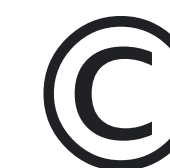
arcu

r. Sed

One option is to use a **Regression Tree**.

Dosage < 14.5

4.2% Effective   **Dosage >= 29**

id

rcu

2.5% Effective   **Dosage >= 23.5**   r. Sed

quam nunc, varius non lectus non, pretium gravida purus.

52.8% Effective   **100% Effective**
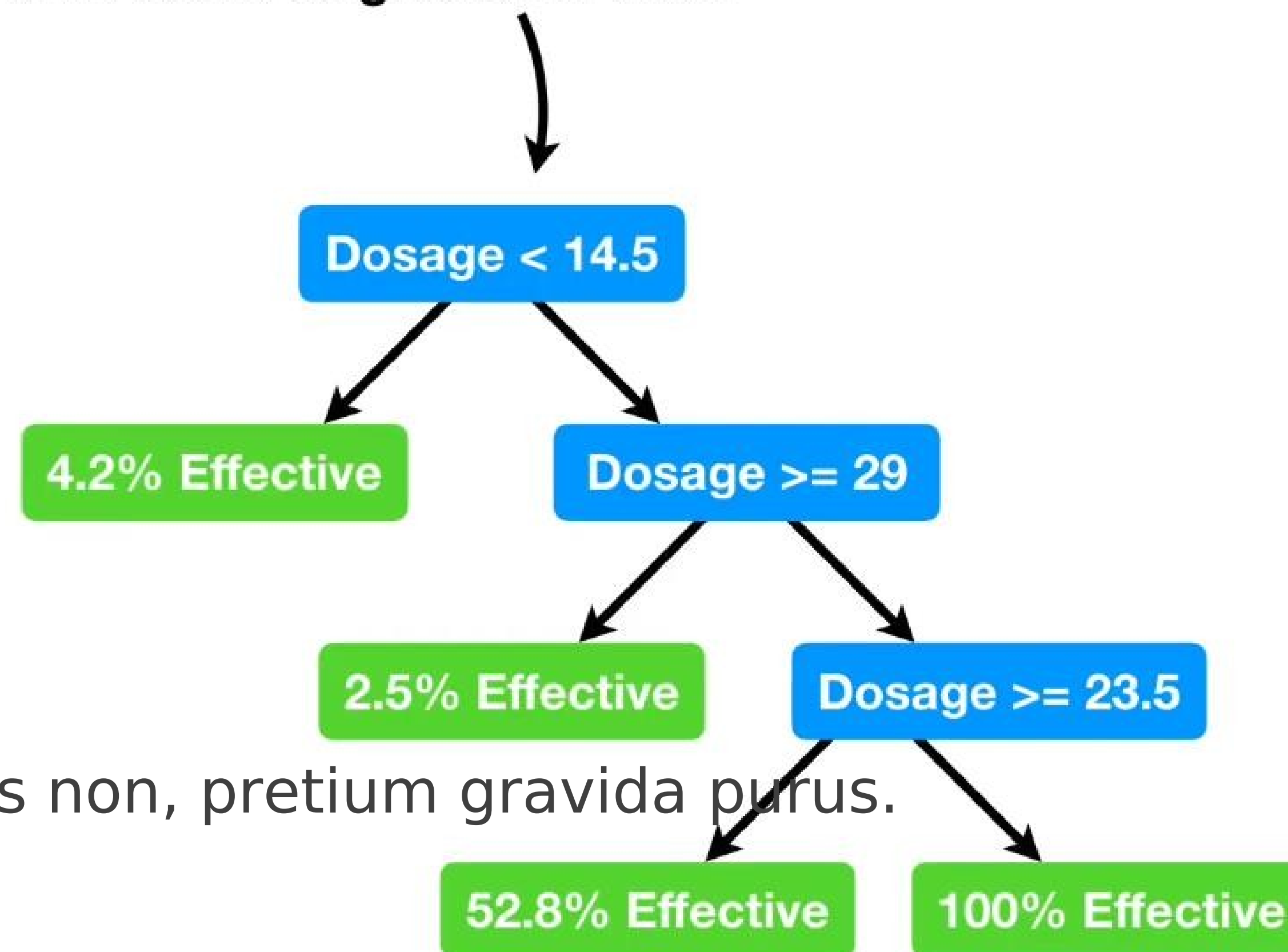
In a **Regression Tree**, each leaf
represents a numeric value.
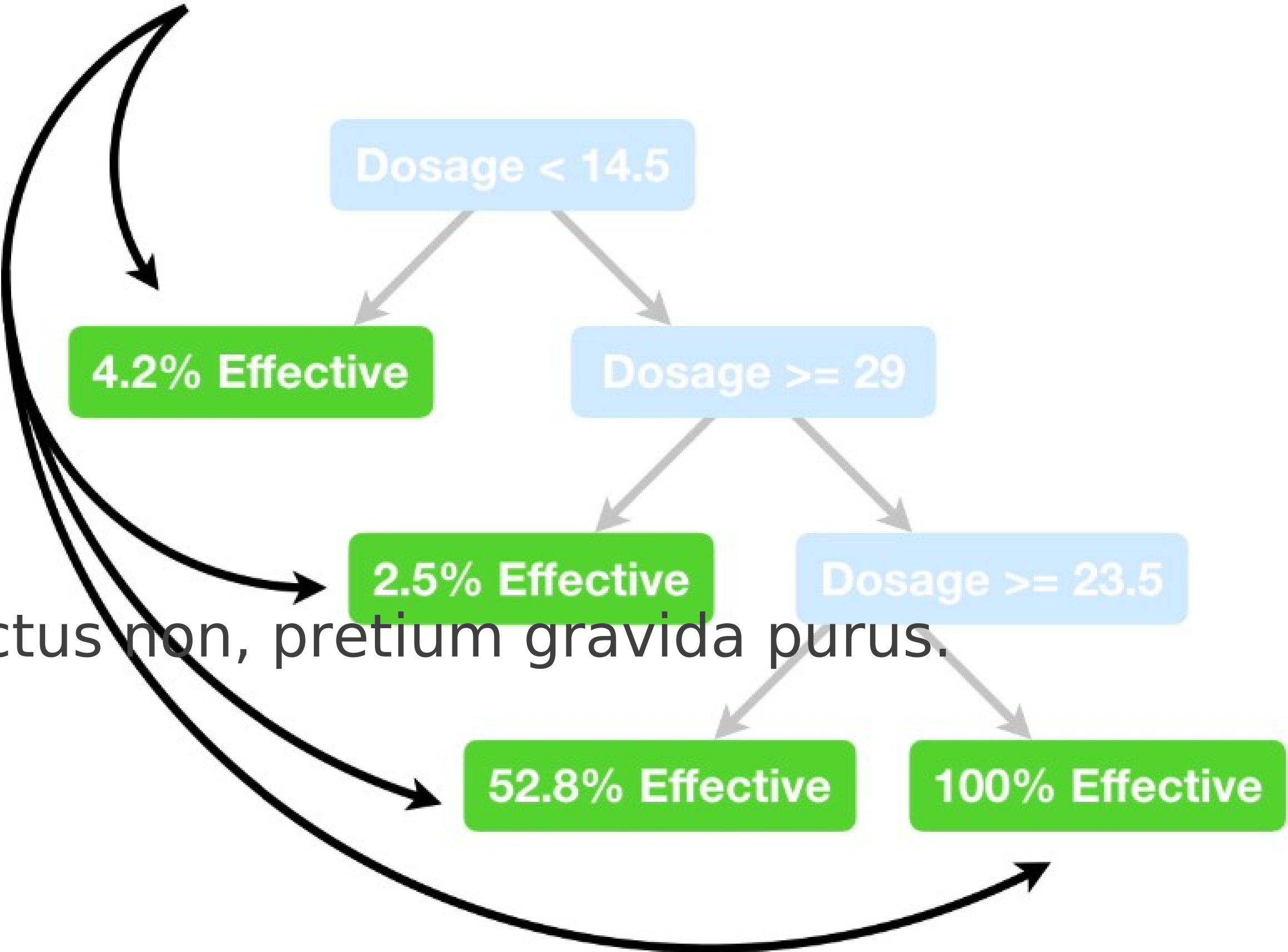
Dosage < 14.5

4.2% Effective

Dosage >= 29

re id

et

r arcu

2.5% Effective

Dosage >= 23.5

tur. Sed

quam nunc, varius non lectus non, pretium gravida purus.

52.8% Effective

100% Effective

Developer Student Clubs

Google Developers

With *this* **Regression Tree**, we start by
asking if the **Dosage** is less than **14.5**.

Dosage < 14.5

4.2% Effective

Dosage >= 29

2.5% Effective

Dosage >= 23.5

52.8% Effective

100% Effective

id

rcu

. Sed

quam nunc, varius non lectus non, pretium gravida purus.

Developer Student Clubs

Google Developers
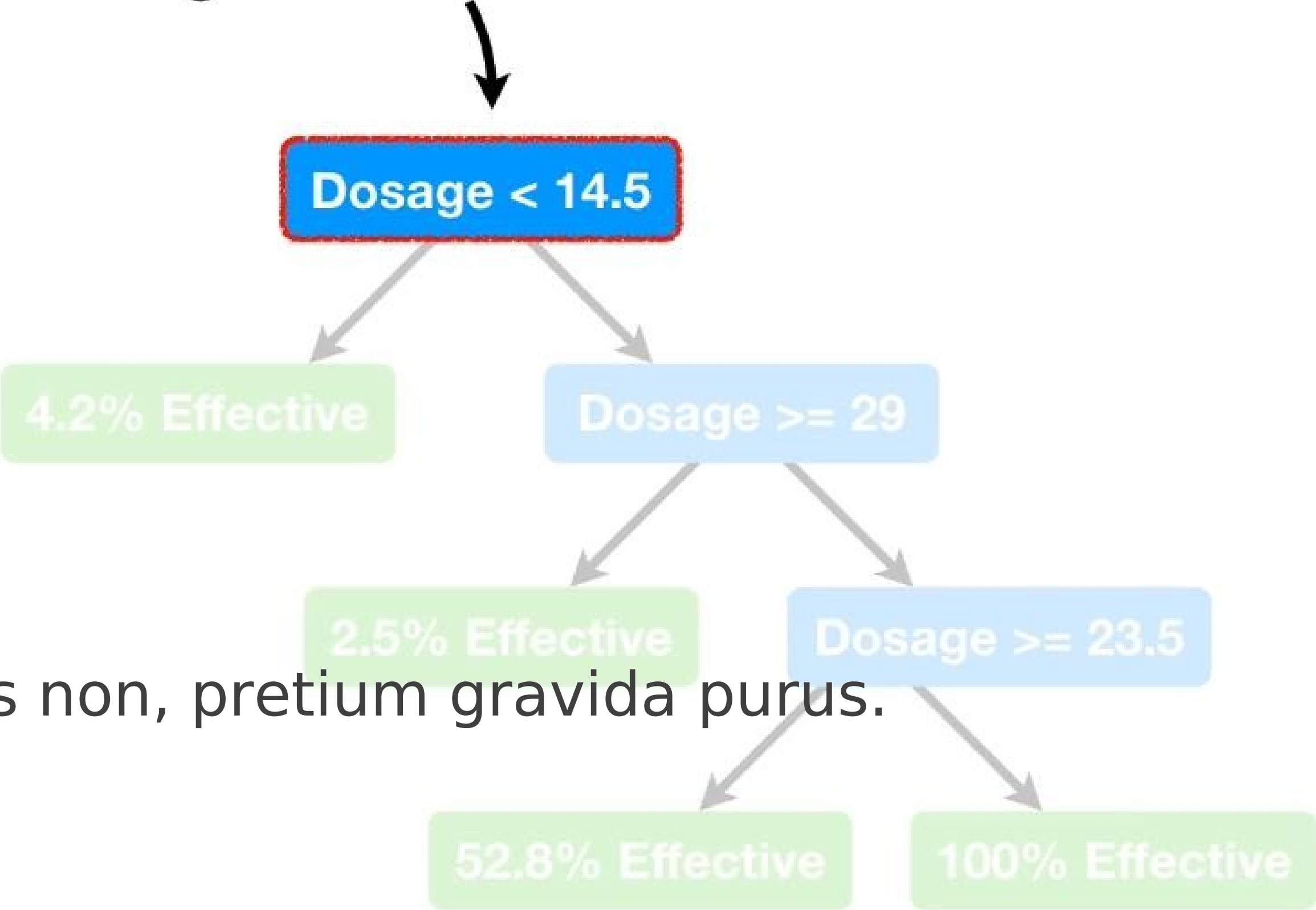
...and the average **Drug Effectiveness** for these **6** observations is **4.2%**...
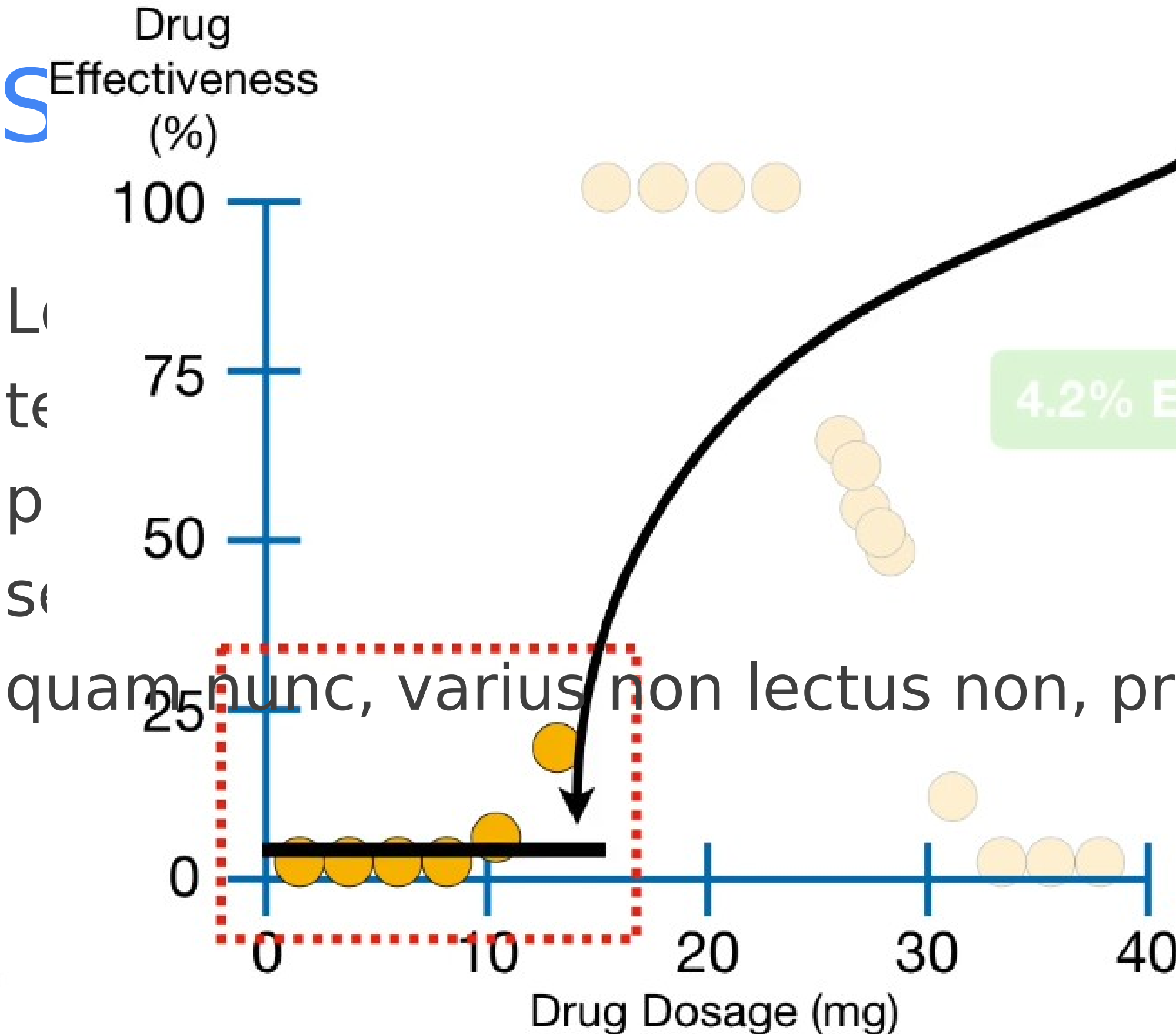
Drug Effectiveness (%)

100

75

50

25

0

0   10   20   30   40

Drug Dosage (mg)

Dosage < 14.5

4.2% Effective

Dosage >= 29

2.5% Effective

Dosage >= 23.5

52.8% Effective

100% Effective

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed quam nunc, varius non lectus non, pretium gravida purus.

...so the tree uses the average value,
52.8%, as its prediction for people with
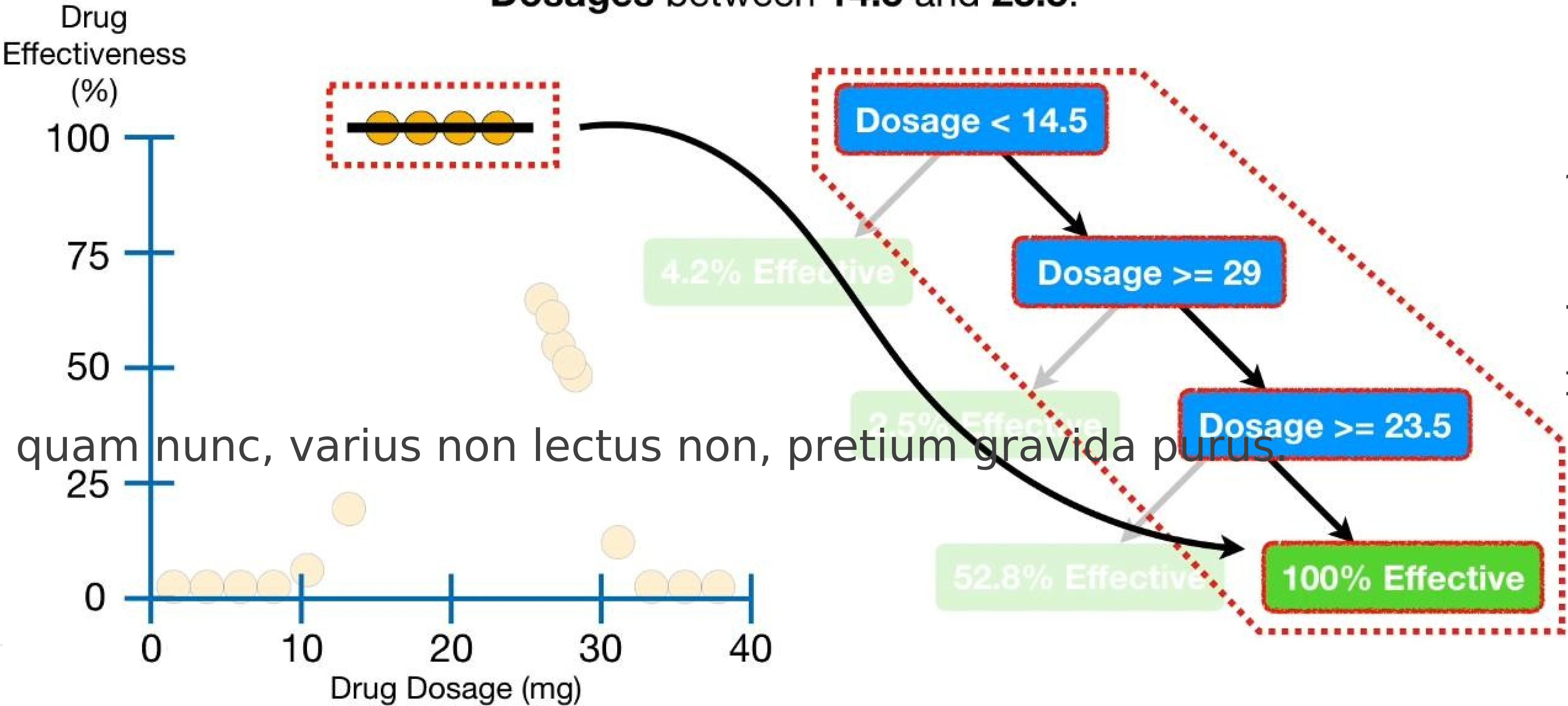**Dosages** between **23.5** and **29**.
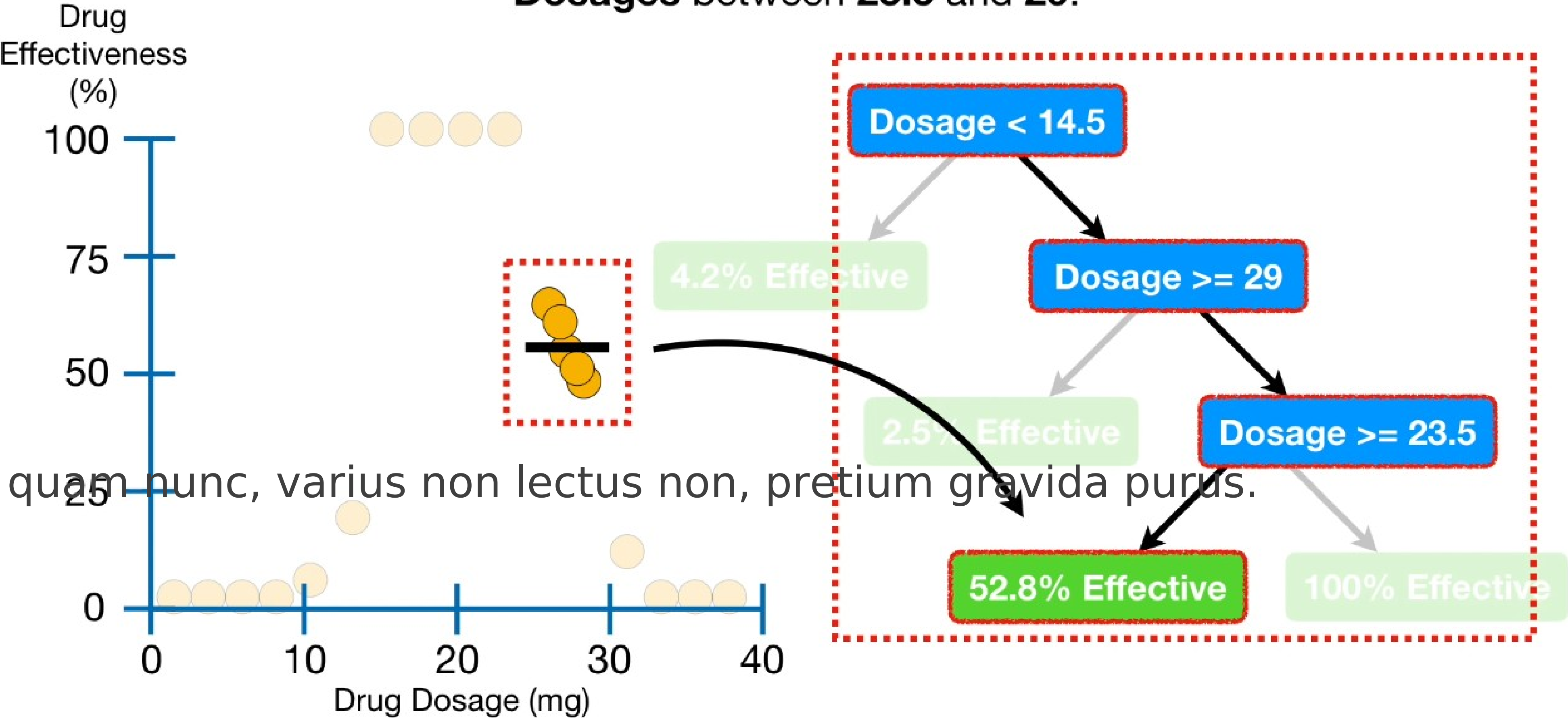
quam nunc, varius non lectus non, pretium gravida purus.

...so the tree uses the average value,
100%, as its prediction for people with
Dosages between **14.5** and **23.5**.

Drug
Effectiveness
(%)

100

75

50

25

0

0   10   20   30   40

Drug Dosage (mg)

Dosage < 14.5

4.2% Effective

Dosage >= 29

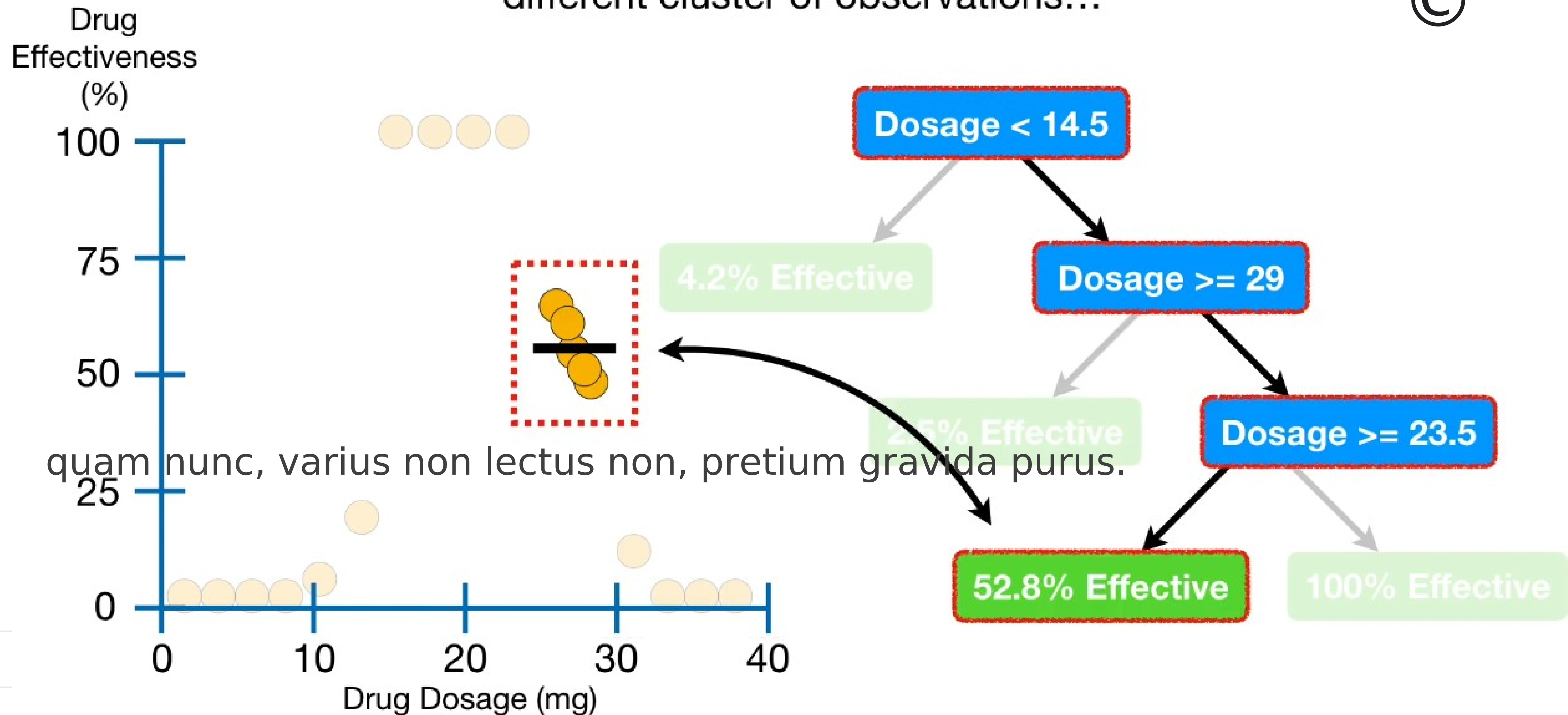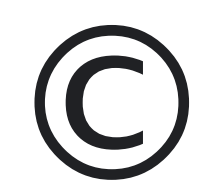Dosage >= 23.5

52.8% Effective

100% Effective

re id
t
arcu
ur. Sed

quam nunc, varius non lectus non, pretium gravida purus.

Since each leaf corresponds to the average **Drug Effectiveness** in a different cluster of observations…

At this point you might be thinking, "The **Regression Tree** is cool, but I can also predict **Drug Effectiveness** just by looking at the graph…"

Drug Effectiveness (%)

quam nunc, varius non lectus non, pretium gravida purus.

Drug Dosage (mg)
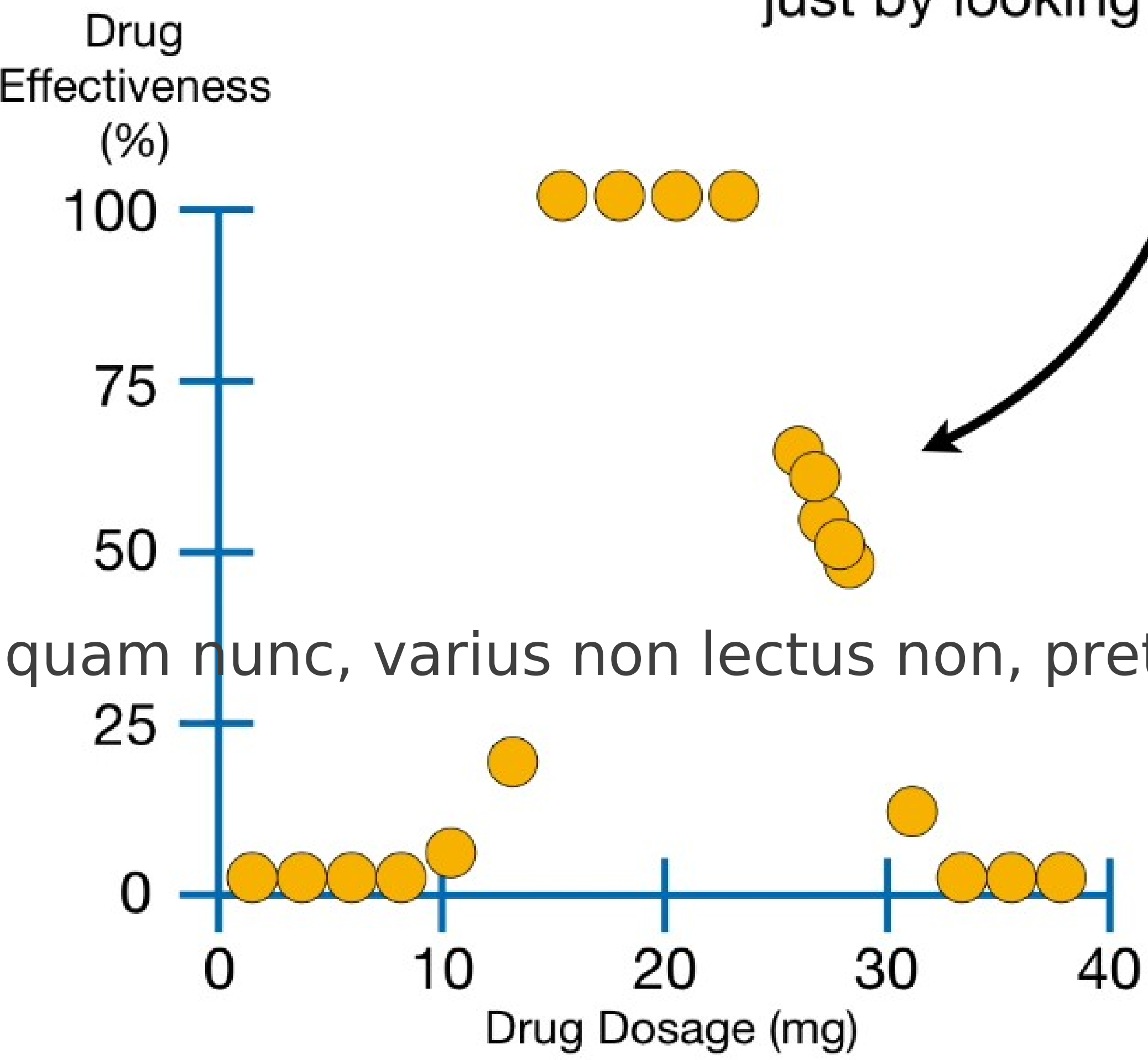
But when we have **3** or more predictors, like **Dosage**, **Age** and **Sex**, to predict **Drug Effectiveness**, drawing a graph is very difficult, if not impossible.
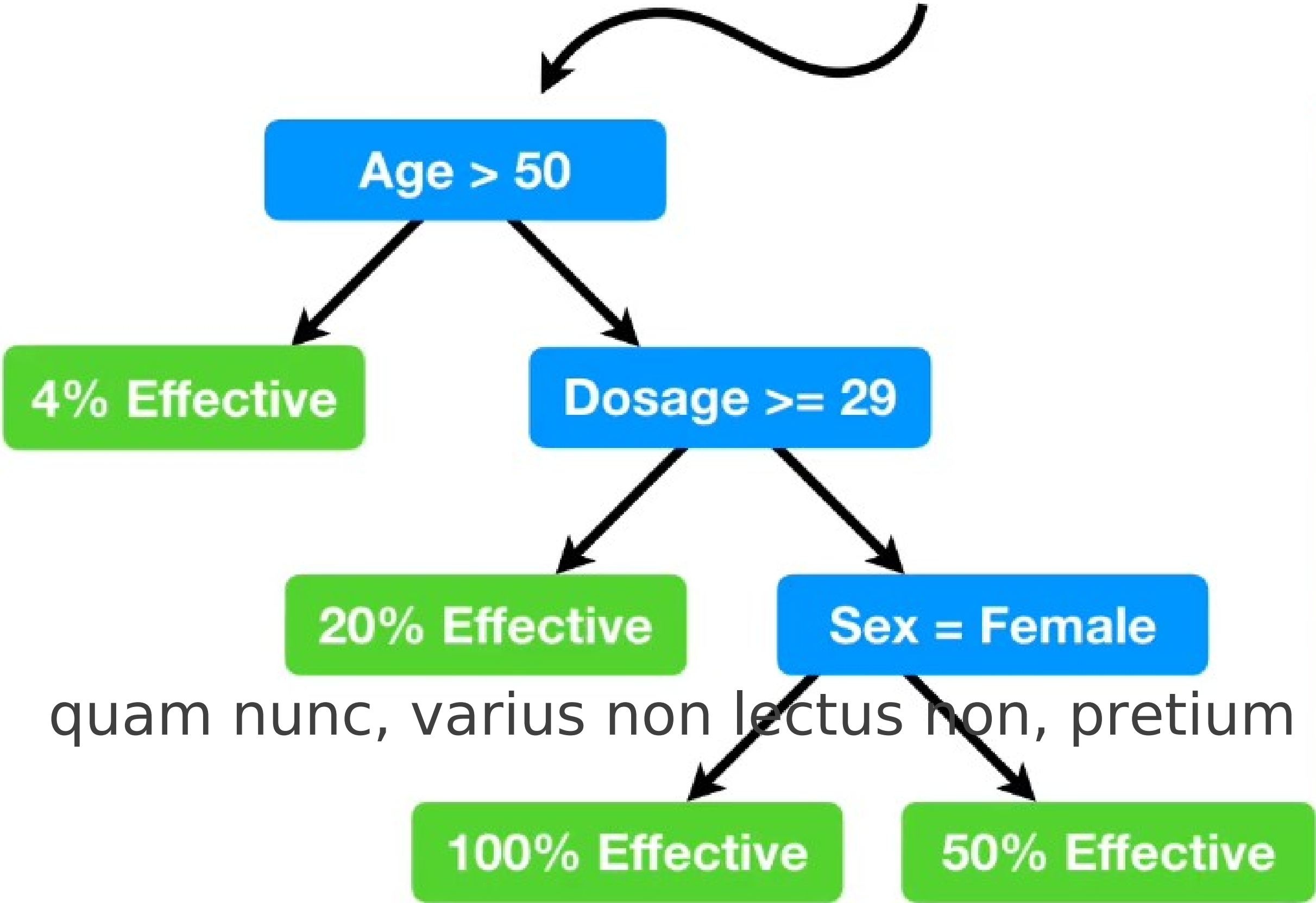
Dosage	Age	Sex	Etc.	Drug Effect.
10	25	Female	…	98
20	73	Male	…	0
35	54	Female	…	100
5	12	Male	…	44
etc…	etc…	etc…	etc…	etc…

quam nunc, varius non lectus non, pretium gravida purus

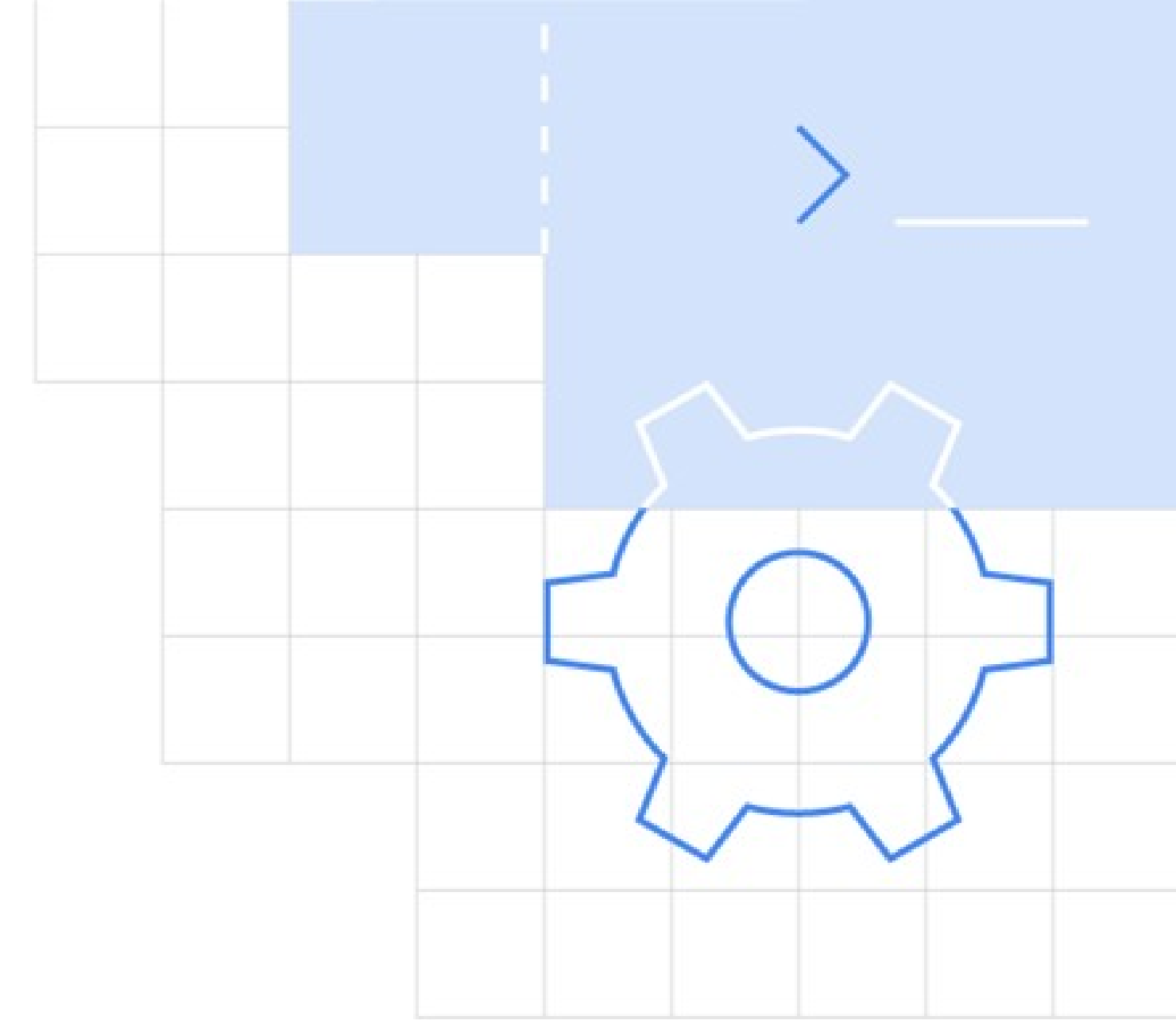In contrast, a **Regression Tree** easily accommodates the additional predictors.

Age > 50

4% Effective

Dosage >= 29

20% Effective

Sex = Female

100% Effective

50% Effective

Dosage	Age	Sex	Etc.	Drug Effect.
10	25	Female	...	98
20	73	Male	...	0
35	54	Female	...	100
5	12	Male	...	44
etc...	etc...	etc...	etc...	etc...

quam nunc, varius non lectus non, pretium gravida purus.

Developer Student Clubs

Google Developers

Now you fully understand the concept

Decision Trees

Random Forest

Developer Student Clubs