Vaccine efficiency Study

Abstract :

In this study we will ask a simple question : have vaccines proven to be effective. What we mean by that is, is there a direct correlation between the use of vaccines on a portion of population and the reduction of the number of cases of illness prevented by said vaccine. We will focus on the Tetanus vaccine, the DTP.

We start here by seeing that simply showing a decrease in diseases accompanying an increase in vaccines over the years, while providing positive insights, does not prove the efficiency of vaccines. We then try to derive another approach to prove actual causality.

Database :

All the data used in this study is extracted from the OMS website, in xls format :
here is an extract of the DTP vaccination coverage percentage in countries.

	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Afghanistan	92	99*	99*	98	99*	94	86	84	85	84	81	80	74	66	62	59	48
Albania	99	99	99	99	99	99	99	99	99	98	98	99	98	97	98	97	98
Algeria	99*	99*	99*	99*	99*	99*	99*	98	97	98	98	94	93	93	93	92	
Angola	97	95	99	97	99*	99*	99*	93	99*	99*	66	62	75	68	65	57	44
Antigua and Barbuda	88	99*	99*	98	90	75	90	99*	99*	99*	96	99	91		98	89	
Argentina	88	94	98	94	94	94	95	94	97	94	93	93	96			92	88
Armenia	97	97	97	97	98	98	98	97	96								
Australia					92	92											
Austria	95							61	65	68	69	91	97	68	77	75	75
Azerbaijan	98	97	96	95	95	96	97	96	97	97	97	95	98	98	97	98	99
Bahamas (the)	95	99*	96	99*	99	99*	99*	99*	99*	99*	95	99	99*			99*	99*
Bahrain	99	98	98	99	99*	99	99*	98	99*	97	99*	99	97	99	99*	99	98
Bangladesh	94	93	93	91	99	99	99	99	98	98	97	95	97	91	86	87	88
Barbados	99*	96	98					93	85	93	81			91		92	93
Belarus	99	99	97	99	99	99	97	98	98	99	99	99					
Belgium	99	99*	99*	99*	99*	99*	99*	99*	99	99	99	98	97	97	97	97	97
Belize	94	97	96	96	96	98	99*	99	98	98	98	95	98	99	99*		88
Benin	76	98	94	98	94	99*	99*	99*	99*	99*	99*	99*	99	99*	99*	99	99*
Bhutan	99	99*	99*	97	97	93	94	96	97	92	98	97	93	98	86	87	99
Bolivia (Plurinational State of)	93	97	94	86	85	90	87	87	87	85	87	93	88	99	99*	97	98
Bosnia and Herzegovina	90	89	92		95	94	89	95	95	96	94	95	93	93	86	94	90
Botswana	99	87	99	99*	99*	99*	99*	99*	99*	99*	99*	99*	98	99*	99	82	94
Brazil	95	97	99*	99*	99*	99*	99*	99*	99*	99*	99*	99*	99*	99*	99*	99*	99*
Brunei Darussalam	99*	99*	99*	99*	99*	96	98	96	97	99	99*	99*	99*			99*	99
Bulgaria	94	93	90	96	96	96	96	96	97	96	97	97	95	97	90	95	95
Burkina Faso	95	95	95	94	94	93	93	99*	99*	99*	99*	99*	99*	99*	89	86	84
Burundi	97	97	98	98	98	99*	99	99*	99*	98	99	96	95	96	99*	70	74
Cabo Verde	96	92	99	93	98	99	99*	73	80	83	74	75	78	89	96	79	93
Cambodia	99*	99*	99*	95	97	96	93	99*	95	87	85	85	92	74	60	65	78
Cameroon	92	92	93	95	94	90	92	88	93	90	87	85	80	79	71	54	54

And here is an extract of the tetanus cases accounted for in (almost) the same countries :

	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Afghanistan	37	74	39	24	37	20	23	19	57	71	43		81	121	975		
Albania				0	0	0	1	0	0	2	2	1	1	0	1	2	2
Algeria	0	0	0	0	0	0	0	3	9	11	9	18	13			20	
Andorra				0	0	0	0	0	0	0	0	0	0	0	0	0	0
Angola		305	330	360	543	953	490	675	601	790	1'633	89	153	643	964		
Antigua and Barbuda	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
Argentina	8	10	8	11	10	1	5	15	8	6	5	14	25	25	18	11	12
Armenia	1	0	1	0	1	0	3			0	2	0	1	0	1	3	1
Australia	7	3	3	4	7	3	2	3	4	3		2	5	3	2	3	7
Austria					0	0	0	0	0	0	0					0	
Azerbaijan	13	8	9	5	7	0	0		6	3	2	2	2	1	0	1	3
Bahamas (the)	0	0	0	0	0	0	0	0	0	1	0		0		0	0	0
Bahrain	0	0	0	0	0	0	0	0	0	0	0		0		0	0	0
Bangladesh	441	559		508	614	644	710	791	943	1'034	1'235	1'388	1'897	715	1'036	1'221	1'155
Barbados	0	0	0	0	0	0	0	2	2	2	0		0		0	0	1
Belarus	0	0		0	0	0	2	1	0	0	0	1	2	2	4	2	1
Belgium					0	0	0	2	2	1	1	3	2	1	1	3	
Belize	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Benin	122	7	3	8	7	6	82	4	7	9		12	17	25			288
Bhutan	1	0	4	0	0	3	0	23	7			1				0	2
Bolivia (Plurinational State of)	0	0	0	0	0	1	0	9	13	9	10	15	29	16	29	3	8
Bosnia and Herzegovina	0	0			0	0	2	0	0	1	4	0	0	0	0	0	0
Botswana	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0
Brazil	224	273	273	213	314	333	308	278	333	281	431	420	463	513	268	129	346
Brunei Darussalam	0	0	0		0	1	0	0	0			0	1	0			0
Bulgaria	4	0	0	1	2	4	2		2	0	4	2	0	2	2	4	3
Burkina Faso	1	0	96	27	1	3	74	7	11	13	9	35	4	16	4	12	
Burundi			3	1	1	2	2	11		14	23	16		40	71		33
Cabo Verde		0	0	0	2	2	0	0	2	0	0	0		2	1	1	2
Cambodia	0		18	13	15			27	324	242	806	68	1'041	796	966	913	295
Cameroon	58	120	110	43	23	93	83	72	65	147		129	154				279

This database has several defects. First, the countries are not exactly the same. Second, the data for vaccination coverage only goes as far back as 2000 while we could potentially use tetanus cases up to 1982. Third, many values are missing. Finally, we have no data below 50% vaccination.

As basic fixes for these first three issues we will use a hand cleaned version containing only the first 30 countries, going only to the year 2000 and ignoring missing values either in vaccination coverage or tetanus cases accounted for.

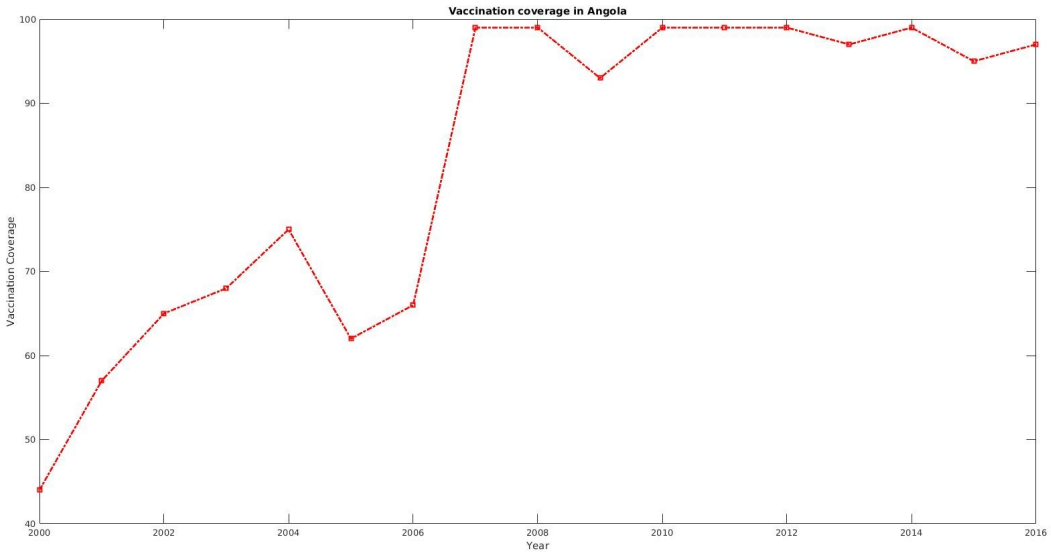
First analysis :

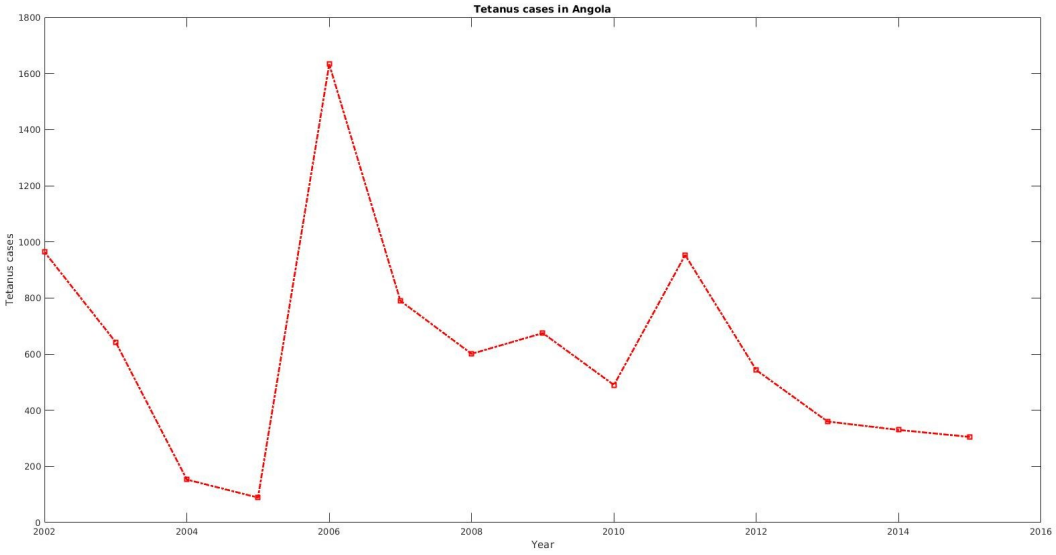
DTP vaccine coverage and Tetanus cases in Angola over the past years.

```
completeTetanosCases = xlsread('../data/tetanosCases.xls');
completeVaccineCoverage = xlsread('../data/DTP1coverage.xls');
```

```
angolaTetanosCases = completeTetanosCases(4,:);
angolaVaccineCoverage = completeVaccineCoverage(4,:);
```

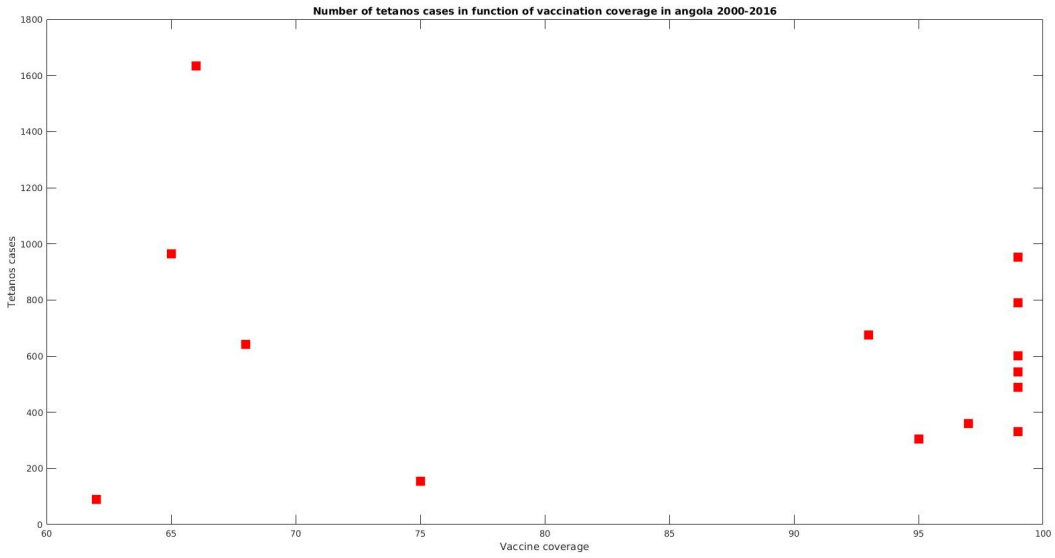
```
plot(2016:-1:2000,angolaTetanosCases,'-.sr','LineWidth',2);
plot(2016:-1:2000,angolaVaccineCoverage,'-.sr','LineWidth',2);
```





To better see if the data is linked, we also derived the number of cases in function of the vaccination coverage. Comparing percentage of coverage and number of cases makes sense only if we assume a somewhat constant population. We will come back on this hypothesis later.

```
plot(angolaVaccineCoverage,angolaTetanosCases,'or');
```



From this kind of limited data we can't conclude on the efficiency of vaccines at all for several reasons :

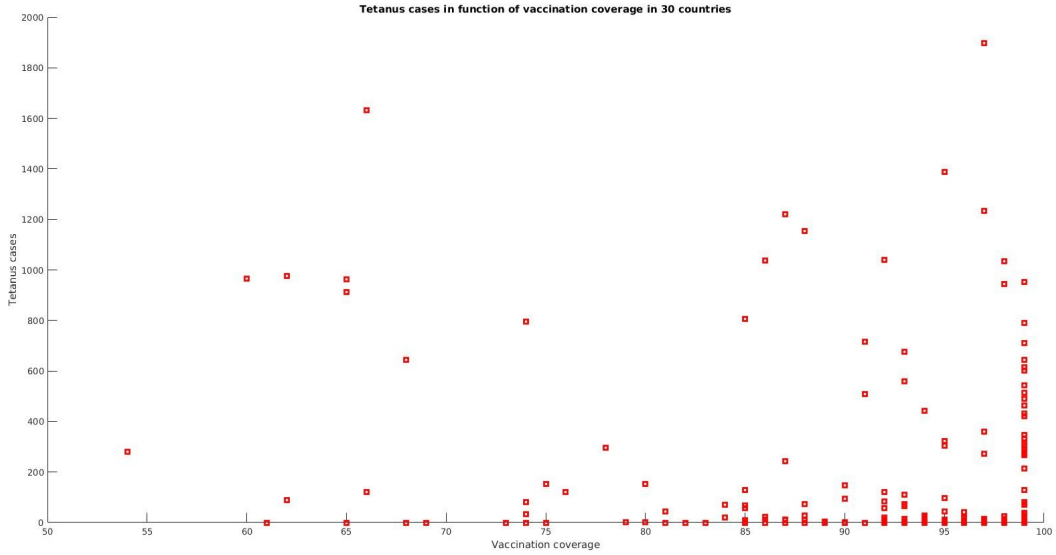
The data is not big enough. We would need at least twice the samples to get a statistically sound tendency. Adding other country as sources could provide statistical soundness.

Second Analysis :

Let's now use the data on DTP vaccine coverage and tetanus cases in the last 17 years in 30 different countries.

We need a way to “average” the data so that it can provide a statistically sound compound. A first idea, visual, is to display the number of cases in function of the vaccination coverage with this time much more data.

```
figure(1);
for i = 1:30
  hold on;
  plot(completeVaccineCoverage(i,:),completeTetanosCases(i,:),'sr','LineWidth',3);
end
```



It seems we still do not get even a visual tendency. Several reasons can cause that (assuming there actually is a relation between vaccination and a reduction of tetanus cases). First, as stated before, the population of many of the country involved is not constant. It is however not varying quickly enough to justify suppressing correlation, so we will keep a study of cases percentage in function of vaccination percentage for a final cleaning of data. Second, there could be many outliers, either because of a sudden outbreak of a disease in a particular country or because of a country never having known the disease.

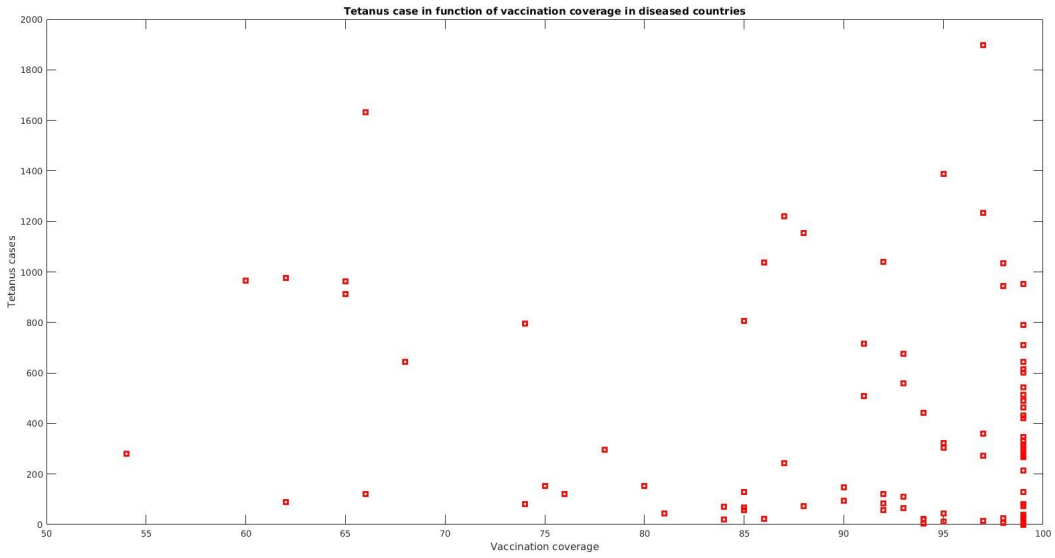
Still we can see a strong pull towards the point of 100% vaccination and 0 cases of disease. This could be explained by the many countries providing data between 90 and 100% coverage with 0 disease in the past years. The presence of a disease can vary for several reasons, and once no cases erupt, the chance for the disease to resurface might be much lower.

Third analysis :

As stated, keeping only the countries that have seen a high number a disease outbreak during the past 17 years might clean the data, as the others may never have needed vaccination and might not be good test subjects.

We only study here the number of tetanus cases in function of the vaccination percent for countries having had at least a year with 100 or more cases (the arbitrary nature of this number should be called into question later).

```
figure(1);
for i = 1:30
  for j = 1:17
    if completeTetanosCases(i,j)>100
      plot(completeVaccineCoverage(i,:),completeTetanosCases(i,:),'sr','LineWidth',3);
      hold on;
    end
  end
end
```



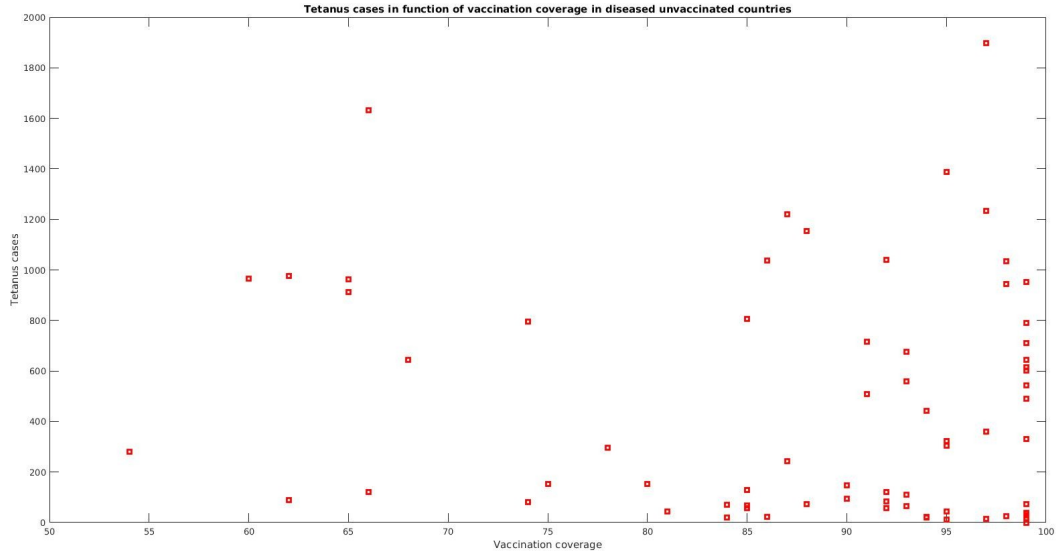
We can see the pull towards the bottom right remains strong, pointing at a correlation between high vaccination coverage and low number of tetanus cases.

We however need to remain aware that this does not prove that the DTP vaccine causes a reduced amount of tetanus cases. They could both be caused by a third factor at the same pace. It could be that the countries in this study have been able to cover more population with the vaccine because of increased budget, which also led to better sanitary services that could be the sole reason for the reduction in tetanus cases.

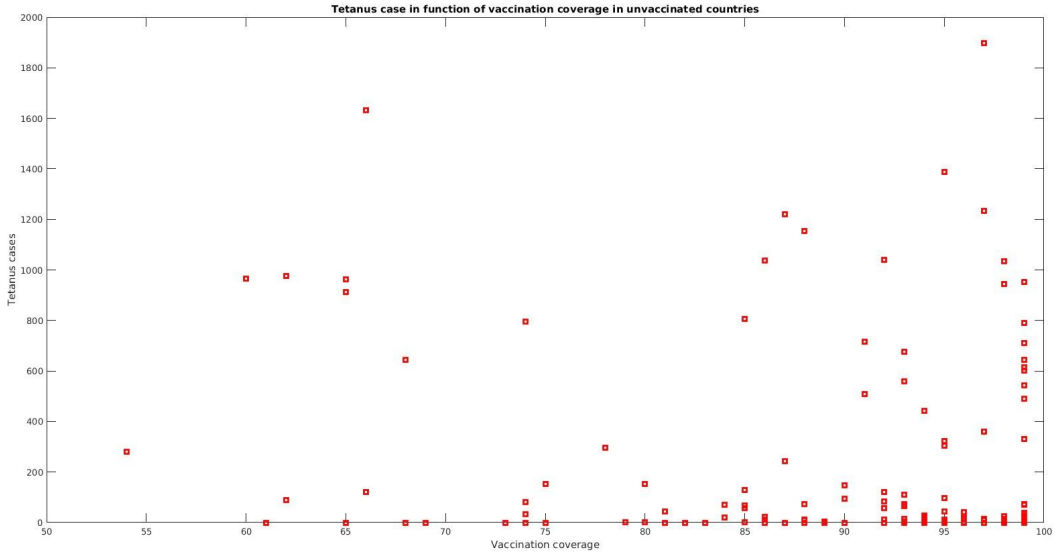
Fourth analysis :

It could be interesting to further clean the data by removing the countries in which there has always been vaccination (at least since 2000) to a very high and unchanging rate.
We consider a country to have always been covered if it had a vaccination coverage higher than 95% in 2000, which is appropriate only because vaccination rate do not tend to drop over time for any country in the database.

```
figure(1);
for i = 1:30
  for j = 1:17
    if completeTetanosCases(i,j)>100
      if completeVaccineCoverage(i,17) < 60
        plot(completeVaccineCoverage(i,:),completeTetanosCases(i,:),'sr','LineWidth',3);
        hold on;
      end
    end
  end
end
```



Here is the version were we remove only vaccinated countries and keep non diseased ones :



We could only reach the conclusion of a mild tendency for the number of cases to drop towards zero when the vaccination rate goes towards 100%. There are several factors making this approach doomed to fail. First is the absence of data below 60% vaccination coverage. Second is the presence of many outliers, even after various attempts at cleaning the data. These could be due to sudden outbreaks of the disease linked to outside factors. The third and main reason is these outside factors. What this means is that these data are taken in a terrible test environment, without control groups or isolation. This implies that we can only ever hope to show correlation and not causality.

In the next part of the study we wanted to try and find a proper test environment for this DTP efficiency study. Ideally we are left with two options :

Finding a study were two groups are formed, one inoculated with the vaccine, the other not, and the two groups are then either exposed to the disease or followed over a long period. The whole thing would have to be done accounting for outside factors by either removing them or exposing both groups to the same environment. This proper statistical study can't exist for obvious reasons.

An other way is the biological study of the actual effects of taking the DTP vaccine on the body in the short and long term. This goes beyond the scope of the present study, so I will just point towards research that has proven the positive effect of the vaccine (along with side effects) : https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5517a1.htm

Conclusion :

Even after getting statistically sound data by using as much as 30 countries over 17 years as input and after cleaning the data to suppress what could be called “obvious outliers”, we can't reached a definitive answer as to whether the DTP vaccines did cause a reduction in tetanus cases. The best we can say is there appears to be some correlation between more vaccines and less diseases.

As long a no further statistical analysis is provided, proving that vaccines do cause a reduced chance of catching a disease can only be done with medical studies of the actual effect the vaccine has on the body. (Such studies have obviously been provided at length and the mandatory vaccines in most country, including France are well proven to be effective). As a side note, such studies have proven painfully hard to find and of consistence provenance and shape, which is a shame, especially when they need to be the proper response to anti-vaxxers of all sorts.

As an added thought, studies have shown vaccines to have a certain probability to render a subject immune to a disease. The effect on the number on cases appearing in a population that is not completely vaccinated can then be derived, and one can see that there is a threshold of vaccination coverage above which diseases have a sufficiently low chance to erupt. This in terms shows that vaccination doesn't only have a high chance to protect oneself but also is a way to get a much higher chance of protection if a big enough part of the population does it.