

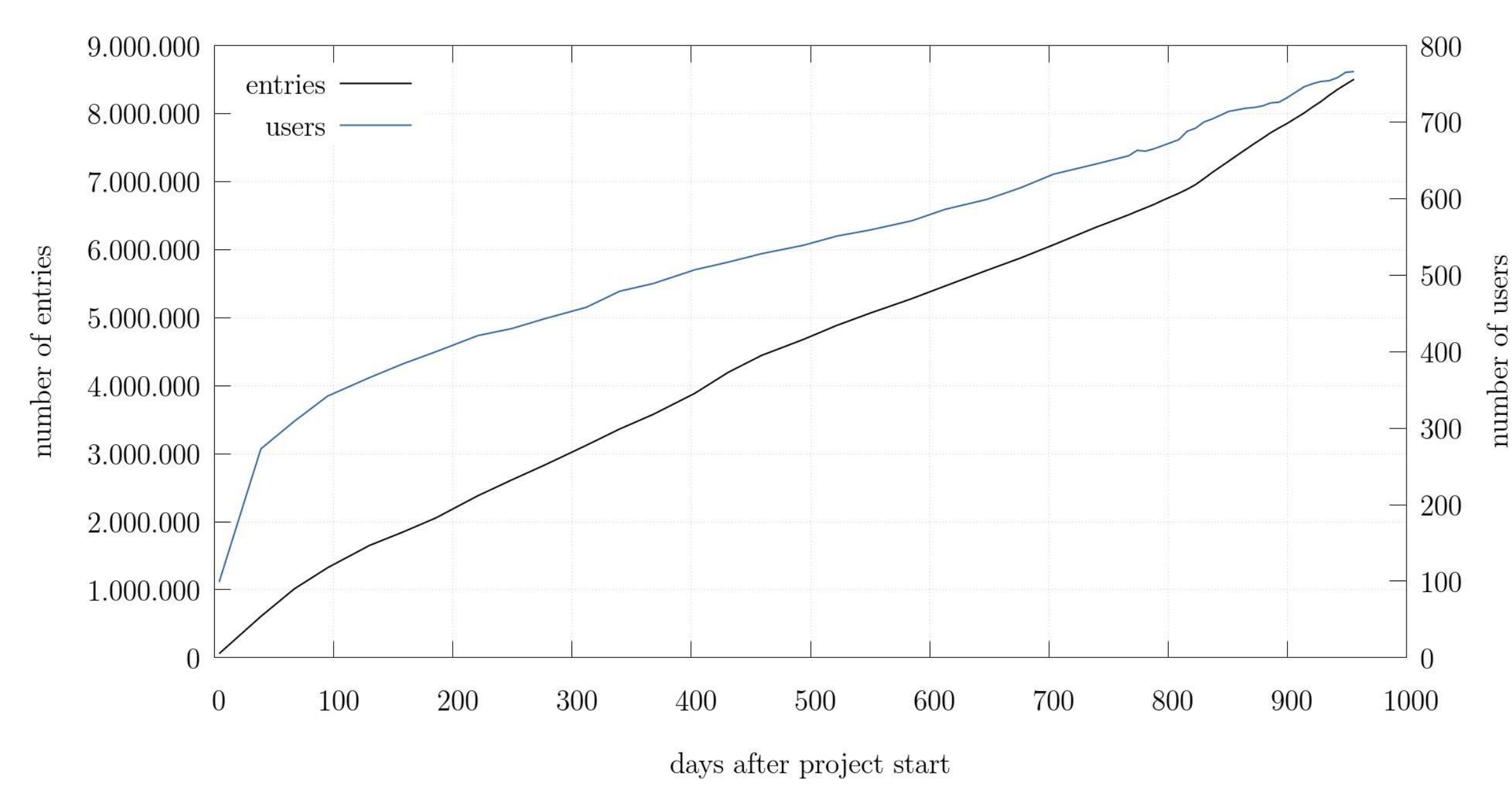
# 750 Volunteers Transcribing 8.5 Million Records Online — an Evaluation

## Deutsche Verlustlisten

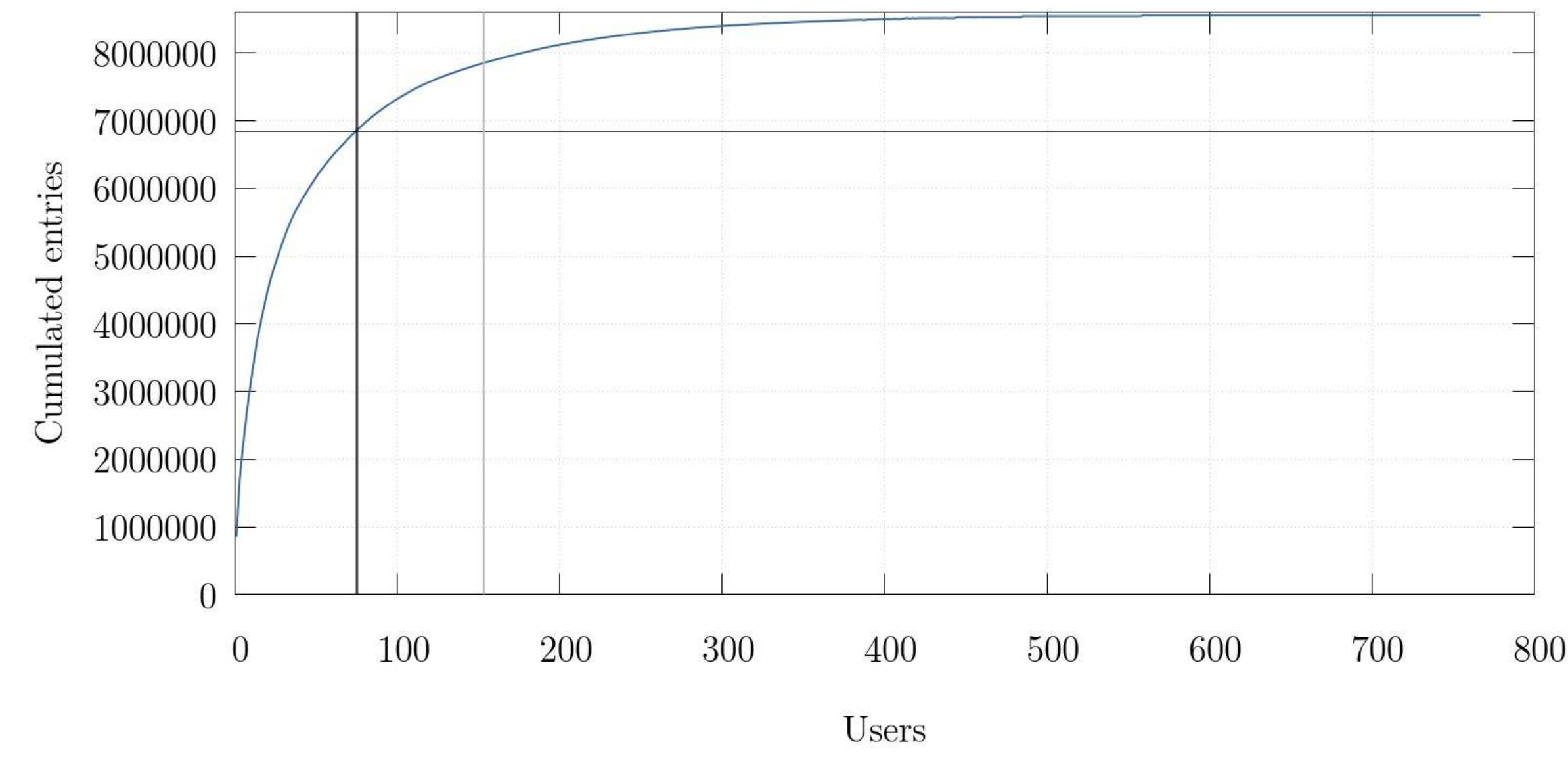


- German WW1 Casualty Lists
- Important source on German WW1 soldiers: official announcements on deaths, woundings, missing, captives
- Ordered by armies and military units
- Published almost daily
- 31,200 pages
- 8,5 million records
- Indexed in 32 months using a novel online crowdsourcing platform

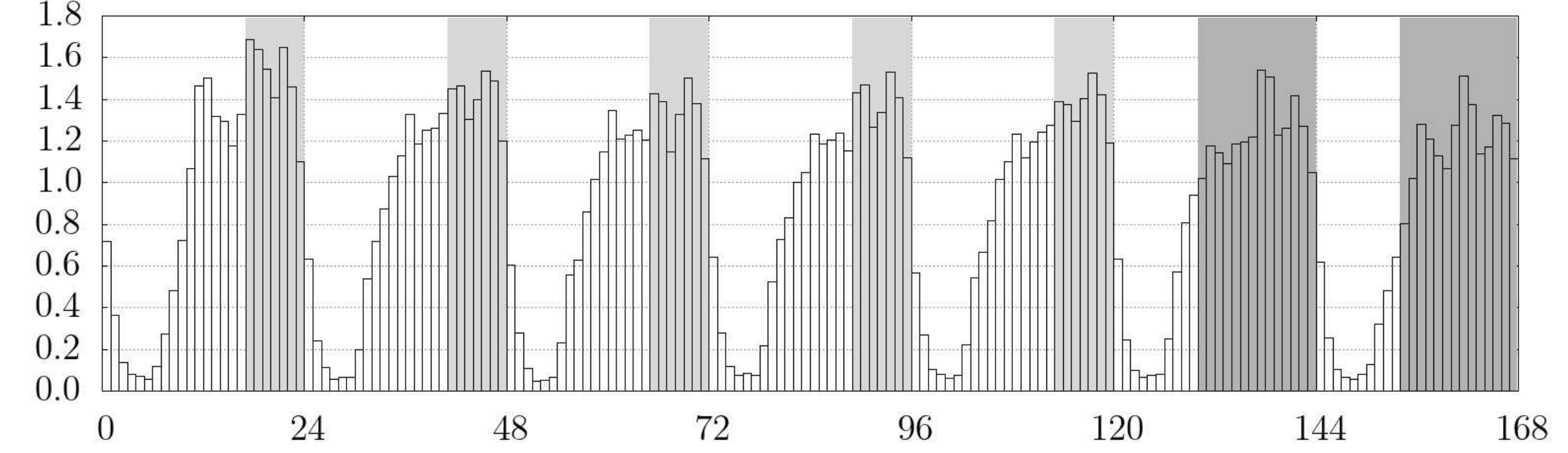
## Progress of the project



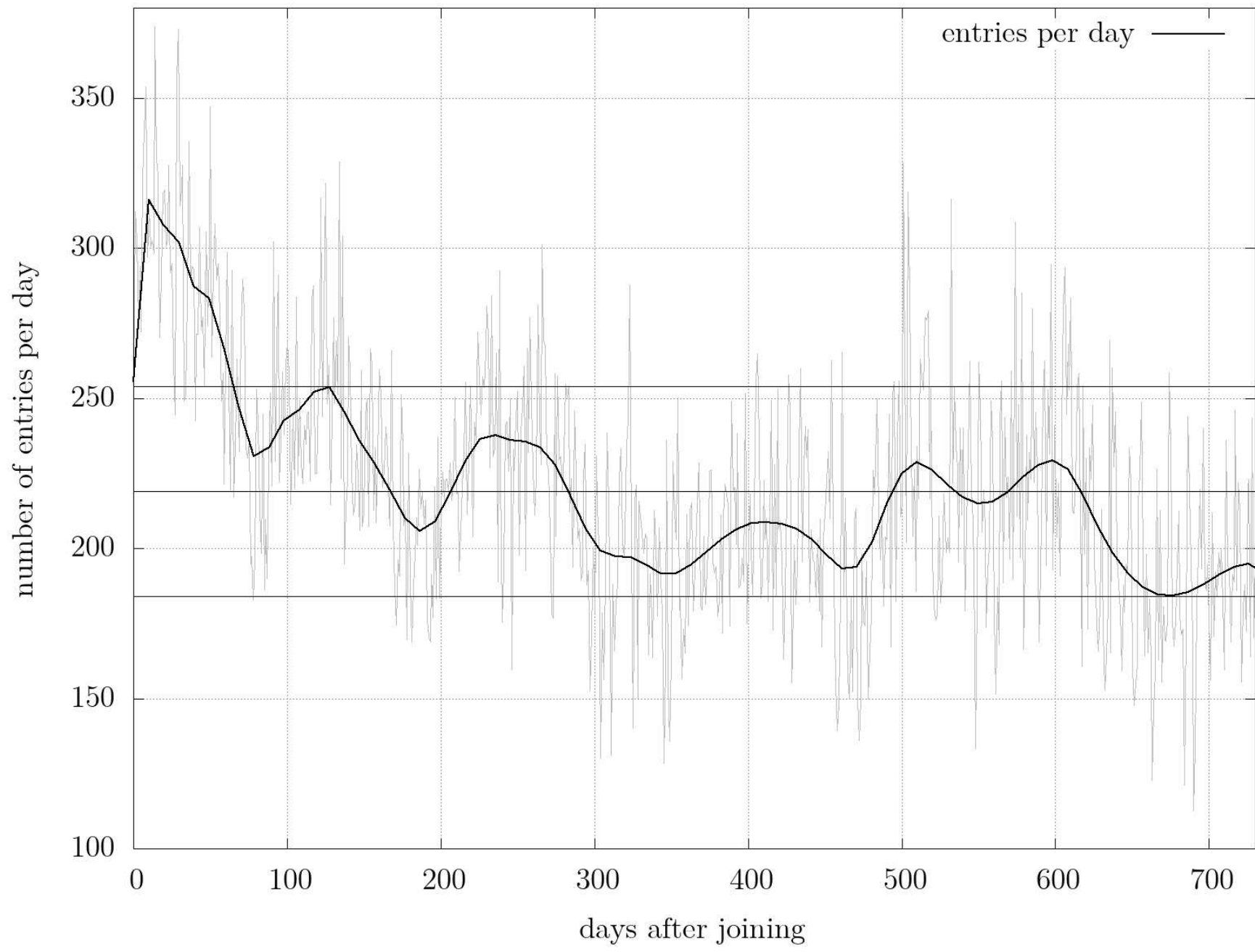
## Hypothesis 1: 80% of the data is entered by 20% of the volunteers. (Pareto principle)



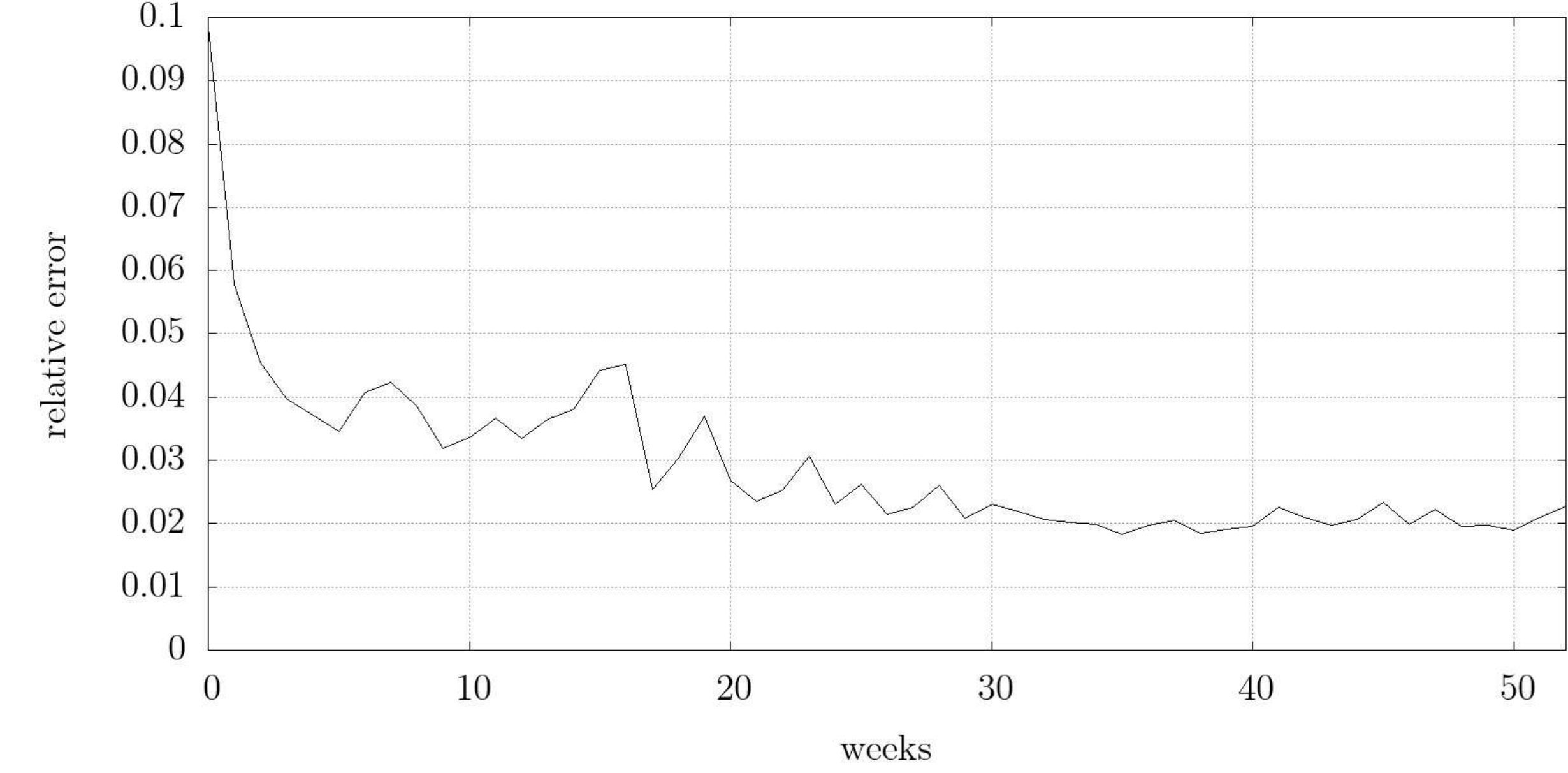
## Hypothesis 2: Most of the work is in the evening hours and on weekends.



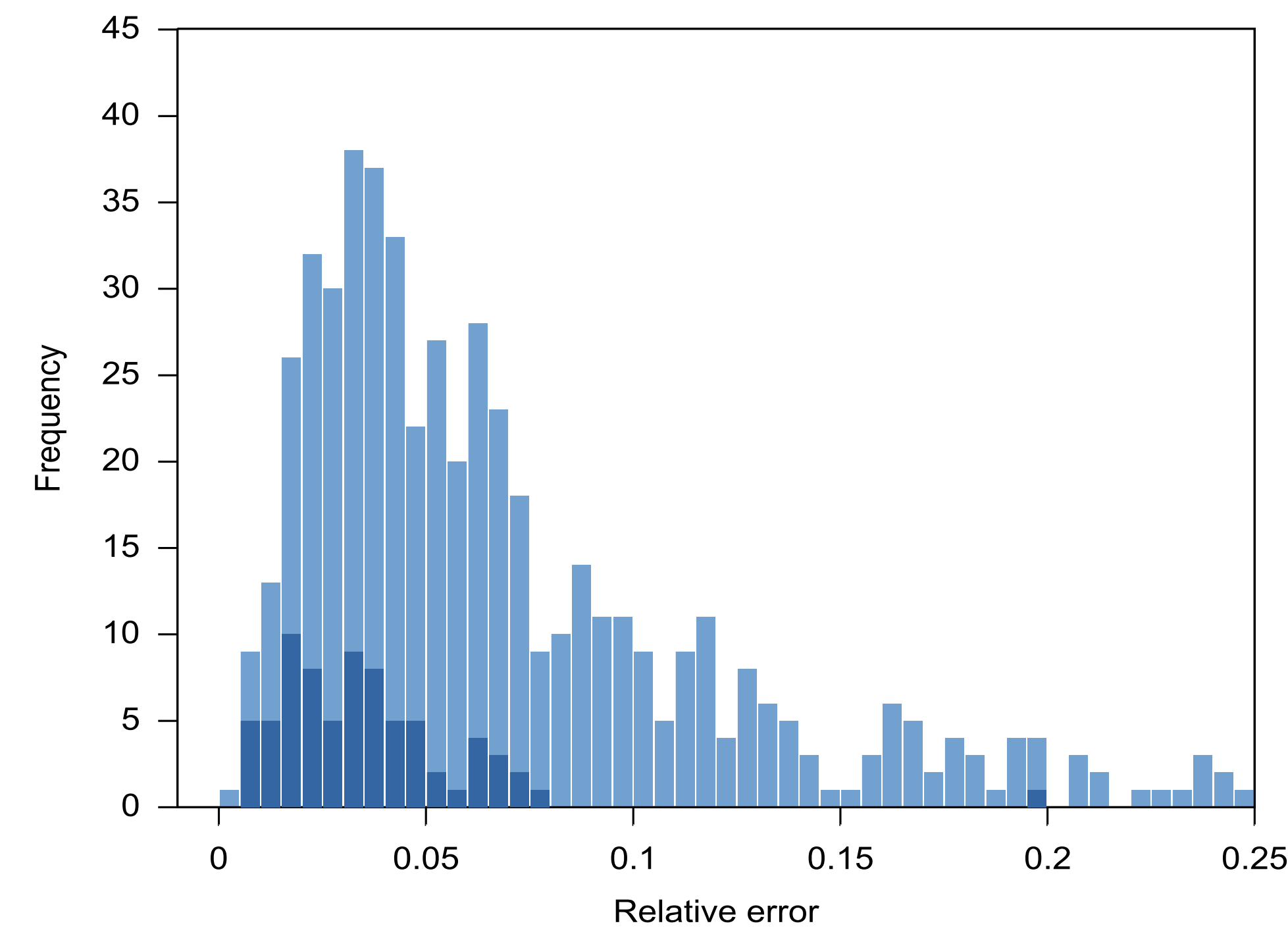
## Hypothesis 4: After a volunteer joins the project the number of edits per day increases strongly, than drops to a long-term average value.



## Hypothesis 5: The longer a volunteer works on the project the less errors he makes.



## Hypothesis 6: Power users make less errors.



"power users" =  
volunteers who entered  
20% of the data

## Hypothesis 7: Volunteers that leave the project after a very short time contribute only little and cause a lot of work.

	count	number of entries	relative error
"power users"	74	78.28%	2.68%
"mayflies"	211	0.32%	19.74%
others	481	19.55%	6.26%