P-4

To be honest at the moment is rather informal. This is not like in traditional software like I have this product definition documents for example. But I have a list of the number of requirements very well sorted, very well defined. Its’ not the case most of the time, more like I want the system is doing this. That could give this type of prediction what can it do for me. Then we come up, we check the data and say this is what we can do for you, this is the base line and performance of the system. We can improve the base line performance. But there is not like a something written.

P-8

I saw plenty of systems, and we still don’t have a good enough methodology for that. Like, these are some checklists that you should go and do. You have some loss function, and this should be the criteria. Well, maybe you should add this and that. No, it’s the discussion is almost never held when it is held, it’s not formal. We know to develop software much better than we know to define nonfunctional requirements for Machine Learning enabled systems.

P-3

Depends on what they are doing. Some are time based, some are based on output, they measure the output and compare against what it should be or some expected or desired value. It’s based on some numeric values, either time based or some other measure that is obtained.

P-5

We do perform interviews and use the result of the system and see how they find the usability. Whenever we deploy the new functionality we always try to follow up some interviews which are also documented and try to follow upon.Some of these ends up in our risk assessment as well and we have a long list of different risks and things that can have consequences for the patients, those are obviously inside the risk management matrix.