

## Questions you might have about (football) stats



2 weeks ago

You can scroll through this page or use the links below to go directly to things you're interested in

- [What is 'per 90'?](#)
- [What is a 'key pass' / What are 'chances created'?](#)
- [What is 'goal contribution'?](#)
- [What is 'sample size' and why do people sometimes say it like it's a problem?](#)
- [What are 'Expected Goals' and what is it used for?](#)
- [Do Expected Goals account for everything?](#)
- [So, what is/isn't in an Expected Goals model \(/what should be\)?](#)
- [What is 'event data' and what is 'tracking data'?](#)

### What is 'per 90'

'Per 90' (or 'p90') = 'per 90 minute chunk'.

Because some players start every game and some come off the bench, it's better to divide stats by a uniform number.

eg: (as of writing) Jesse Lingard has 8 Premier League goals in 17 starts and 11 sub appearances while Dele Alli has 8 goals in 30 starts and 1 sub app, which is awkward to compare properly.

Lingard's played 1556 minutes while Alli's played 2633. You could work out how many minutes they take to score (194.5 vs 329.13) but because a match is 90 minutes long that's the convention.

1556 minutes = 17.29 lots of 90 (or '90s') for a scoring rate of 0.46 per 90; 2633 minutes = 29.26 90s for a scoring rate of 0.27 per 90.

## What is a 'key pass' / What are 'chances created'?

A key pass is the pass which comes before a shot, kind of like a 'shot assist'.

'Chances created' is basically the same thing. To my knowledge, the same stat used to be called 'key passes' but was changed, although different data companies might have different names for similar statistics.

## What is 'goal contribution'?

Goal contribution is the number of goals and assists a player has scored/made.

Sometimes people say a player has been 'directly involved in' a certain number of goals, or 'goals plus assists' (G+A), which are the same thing.

The terms came about as a way of trying to talk about a player's contribution to the team in a shorter way than 'this player scored 8 goals and made 4 assists'. It can also be used to compare players who play different roles, or the same player who has played different roles in different seasons.

eg Dele Alli in 2016/17 scored 18 and set up 7. In 2017/18 (at time of writing) he's scored 8 and set up 10.

His goal contribution, or Goals plus Assists, would be 25 last season and 18 this season. (We could also 'per 90' these stats to give 0.74 Goals plus Assists per 90 (G+A p90) for 16/17 and 0.62 G+A p90 for 17/18)

## What is 'sample size' and why do people sometimes say it like it's a problem?

Sample size is basically how much data you have – a large sample size means you have a large sample of data to work from.

(I guess it's a 'sample' because you'll never have ALL of the data that exists)

If you don't have much data, it's hard to be sure whether the patterns in it will stay the same or change over time. I once flipped a coin 20 times in a row (no word of a lie, I once actually did this out of curiosity) and got heads 13 times.

Now, I *know* that flipping a coin is supposed to be around 50/50 chance, but until I flip the coin another 20 times (or preferably more) I don't know if I've got a dodgy coin. I genuinely did flip the coin another 20 times, this time got heads 9 times.

With football, the average rate that shots are scored is around 10% (though it depends a lot on the type of shot). If I see that a player has scored from 5 of their first 30 shots of the season – 16.6% – I don't know for sure whether this player is better than the average or not. With more data, we can start to be more sure that the patterns we see are legit.

## What are 'Expected Goals' and what is it used for?

Expected Goals (or 'xG') are a way of measuring how good chances are.

Things like place on the pitch the shot was taken and situation leading up to the shot are taken account of in the statistical models that are made to calculate xG.

Each shot gets an xG value based on the results of thousands and thousands of previous examples. This is useful for seeing whether teams are scoring more or fewer goals than we might expect them to, based on how many goals were scored from similar situations over the history of football (that has been coded by data companies like Opta).

But there's a warning about sample size – just like with other statistics, it's safer to draw conclusions with a larger amount of data.

## Do Expected Goals account for everything?

No, Expected Goals models don't account for *everything* for two reasons.

The first is that they can't, because some things (like the way the ball is bouncing) aren't captured in the data.

However, even if *everything* had been recorded for as long as football had been played, even down to the number of decibels the crowd was making, Expected Goals models wouldn't necessarily want to include everything.

'Expected Goals' is perhaps a misleading name, because often xG is used to look at how good a team is at creating good chances.

For example, imagine you could record how cleanly a player struck the ball. You could measure a striker slicing the ball from three yards out and completely missing the target.

If you included how cleanly they struck the ball (ie terribly) the ‘Expected Goal’ value would be very low, but the quality of chance that was created might have actually been very good (if the striker had struck the ball more cleanly).

On top of ‘Expected Goals’ being a potentially confusing name, there are also different types of Expected Goals models, and *some* take into account where the ball ended up (ie a shot off target would have an Expected Goals value of 0 in these models, regardless of how good the chance had been).

## So, what is/isn’t in an Expected Goals model (/what should be)?

- **Place on the pitch shot was taken:** YES
- **Header or feet:** YES
- **Counter-attack:** YES (either by the people who are manually collecting the stats while watching the game saying ‘yes, this is a counter-attack’ or by using the stats available to create a definition of a counter)
- **Position of defenders:** NO, MOSTLY (Exact position of defenders tends not to be collected. Stratagem, for example, collect the number of defenders between a shot and goal, and they and Opta both collect the level of defensive pressure on a shot)
- **Height of the ball** (eg, is the player having to kick the ball at head height): NO
- **What foot the shot is taken with:** MAYBE (Opta, for example, collect what foot shots are taken with. Whether there’s a database of which foot is a player’s stronger foot to compare that to, I’m not sure)
- **Position of the goalkeeper:** SOMETIMES (Opta, for example, collect the position of the goalkeeper for shots on target, either at the point that they save the shot or the point that the ball reaches the goal-line)
- **Speed of the ball** (a fast pass is going to be harder to strike cleanly): NO
- **Quality of shot:** SOMETIMES (While information about what part of the goal (or stands behind the goal) a shot ended up going towards is available, some models are just interested in what happens before the shot. This is to measure ‘quality of the chance’)

There is a caveat to some of these, and that is **tracking data**

## What is ‘event data’ and what is ‘tracking data’?

‘Event’ data is the type you see almost everywhere – passes, tackles, shots.

‘Tracking’ data is where people run (and where the ball goes).

Opta, one of the bigger data collecting companies who provide the stats for places like WhoScored and Squawka, collect event data.

This is done by people watching the games and ‘coding’ what happens by pressing keys on special keyboards. That sounds lo-fi, but they’re very well trained. There’s a short, 90-second video [here](#) (a number of years old now) that gives some idea of the process.

Tracking data can be done in one of two ways: cameras or GPS.

Every now and then you hear a story about a player who’s given their shirt to someone in the crowd, only to have to track them down because the GPS kit was still in it. This is one way tracking data can be created.

Other systems have multiple cameras set up around the stadium to track the players. There’s a 2min30sec video [here](#) from a few years ago about how the system in the Bundesliga works.

The advantage of tracking data is that it can capture where everyone is on the pitch and how fast everything is moving and can be automated, rather than relying on human ‘coders’ to be consistent.

The disadvantage is that it is far more difficult to work with, as it creates a HELL of a lot more data, and the technology behind it can be expensive. It also has the drawback that it doesn’t tell you where a player is facing or what foot they kick with.

However, there is technology being developed to generate stats from TV pictures automatically, which will undoubtedly have pros and cons all of its own.

Categories: [Uncategorized](#)

[Leave a Comment](#)