

FINAL REPORT

*Hanspeter Stamm & Markus Lamprecht
L&S Social Research und Consultation AG Zurich
February 2001*

Overview

Football is without doubt one of the world's most important sports. While there is a relatively large amount of information available about spectator interest in international and national football competitions, there is a lack of reliable data about how many people actually play football worldwide, or are involved in the game as referees, managers, coaches or officials.

With this in mind, in the summer of 2000, FIFA initiated an enquiry under the title *The Big Count*, aimed at collecting and evaluating relevant data. A short questionnaire was sent to every FIFA member association, and the information contained in the replies was electronically stored and then evaluated. Since the first replies received indicated that the data was incomplete in a number of cases or seemed rather implausible, it was necessary to carry out various checks and to correct or complete the information in a number of instances.

This paper describes the procedure used in evaluating and completing the data and presents the overall results obtained from the *Big Count*. At the end of the report is a technical appendix containing details of the techniques used for the estimation procedures involved. The whole data collection – a considerable amount - is contained in a separate document, from which details about individual associations can be obtained.

Procedure

The following information was to be entered by the associations in the questionnaire, categorised according to gender, with estimates being acceptable in the case of a lack of exact figures.

- Number of professional footballers, separated under adults and juniors
- Number of registered footballers, separated under adults and juniors
- Number of non-registered footballers
- Number of referees
- Number of officials
- Number of football clubs
- Number of teams

After the questionnaires had been returned to FIFA, the information was stored in an electronic data bank. For two reasons it was not possible simply to obtain a definite publishable total:

1. Out of the current total of 204 member associations of FIFA, 161 returned the questionnaire (78.9%). At just under 80% this is a reasonable proportion of replies, and the figure is well above what would normally be obtained in the social sciences, but nonetheless, for a complete evaluation of the situation, data from the all national associations was essential. But there were gaps in the data, which had to be filled by appropriate means in order to obtain „real“ overall figures.
2. Looking through the data shows that some associations had had trouble in obtaining all the information requested and so had left certain areas blank.
 - In many cases this was simply due to the fact the figures needed were simply not available. For example, in deciding what is the number of non-registered players, male and female, in most instances the answers will be estimates, which may or may not be very accurate. The estimates received in answer to this question were very disparate: one association classified 46% of its population as non-registered players, while a number of others had the percentage between 7 and 28. A large number of associations were not willing to deal with estimates and either left the space blank or entered an extremely low figure.
 - In addition there seemed to be some miss understanding as to what information was required. This led in some cases to the responses being incomplete or incorrect. For example, it is hard to see why Argentina with a population of some 37 million would only have about 11,000 registered male footballers while its two considerably smaller neighbours, Chile and Uruguay, record their totals as 520,000 and 108,000 respectively, and a recent survey about sports activities in Argentina indicates the proportion involved in organised football there to be much higher.

These examples show that the replies from the associations were not always complete or plausible. For this reason, it was necessary to make minor corrections to a number of entries. Therefore it had to be decided how to make estimates for missing data and how to identify and adjust incorrect or implausible responses.

In both cases the same basic technique was used which is described in detail in the appendix: by using multiple regression estimates, first a statistical model was sought which reflected the available data as accurately as possible. The regression coefficients obtained were then used both to provide estimates of missing data and also to identify possibly false or implausible replies, and to indicate in which direction these needed to be corrected. The correction was based on the one hand on the regression results and on the other on a comparison with the figures from similar associations.

With reference to the number of non-registered players however, this procedure had to be complemented by an extra step in the working. As mentioned above, the data submitted by the associations varied considerably – if responses were made at all. While in a few associations probably „anybody who ever kicked a ball at any time“ was included, most associations seemed to have based their answers on a more or less regular involvement in a footballing activity. In addition it seemed that most associations had not considered the involvement of keen football-playing children. With this as the background, scenarios to reflect the position of football in various societies had to be constructed in order to estimate the numbers of non-registered children and „occasional“ players (see appendix). These figures are listed separately in the following tables under „Children and occasional players“ in order to distinguish this extracted estimate from the figures for more regular players who were recorded by most of the associations.

Results

Tables 1 and 2 contain the results of the above-mentioned control and correction procedures. The presentation shows three differently-obtained types of results:

1. *Original values:* These are the totals of the original responses to the questionnaires of the *Big Count*. As already mentioned, replies from about 40 (mostly smaller) associations were missing.
2. *Informed estimate:* This column contains original and estimated values, with the original values only being replaced by estimates if the originals seemed implausible or were totally missing. The corrected values were not simply obtained from the regressions, but derived on the basis of comparison with similar associations, with the regression estimates serving generally to provide limits for the corrections applied. Corrections were applied particularly to figures from the CONMEBOL region (responses too low) as well as to those from AFC and CAF (data missing). The corrected values as well as the original figures are contained in a table in a special document.
3. *Original values plus estimates:* This column shows the totals of the original values, plus the informed estimates (see column 2) in cases where data was missing. In other words, estimates have only been used for those associations for which data was incomplete; for the 161 associations who replied to the *Big Count* their responses have been entered as received. Thus the entries here are estimates, which, independent of the data checks, are based on the assumption that the information provided by those associations that replied was accurate.

From Table 1 it will first be noted that the figures for the original values and for the informed estimates are very close together. That there is not a greater discrepancy is due, on the one hand, to the fact that most of the larger associations submitted seemingly reliable data. Therefore it was only necessary to correct the figures in a very few cases. On the other hand the estimates for the missing associations are of less significance here since they concern almost exclusively the smaller associations. To be specific, the data that was actually received covers over 90% of the world's population (appx. 5.5 billion based on a total population taken as 6 billion). That the combined „politically correct“ values for men in the final column are higher than the original and the informed estimates, is due to the fact that some associations clearly overestimated the numbers of non-registered players – a problem that was corrected in the informed estimates.

Overall, however, the values are relatively stable. It can be concluded that currently some 220 million men and 20 million women play football worldwide. In terms of total population this represents a proportion of around 4%, or in other words, one adult in 25 plays football at least on a casual basis. If we add to this total of players the number of people who are involved in the game as referees or officials, then the worldwide total comes to about 250 million who are directly involved in football in one way or another.¹

¹ *In this connection it should be pointed out that there are certain groups of people who are not included, yet who are dependent on football. For example, there are the media representatives (journalists, television teams, etc), marketing specialists, producers and distributors of balls, shoes, clothing, as well as those involved in building and maintaining stadiums.*

Table 1: Different estimates based on data obtained from the *Big Count* figures (the numbers are in thousands)

	Original values	Informed estimate	Original values plus estimates
Male registered (incl. professional)†	12'044	12'487	12'456
Male youth	15'114	15'793	15'572
<i>Male registered plus youth (sum)</i>	<i>27'158</i>	<i>28'280</i>	<i>28'028</i>
Male not registered	72'452	82'214	89'046
Male children and occasional*	110'000	110'000	110'000
<i>Male not registered (sum)</i>	<i>182'452</i>	<i>192'214</i>	<i>199'046</i>
<i>Male total (sum)*</i>	<i>209'610</i>	<i>220'494</i>	<i>227'077</i>
Female registered (incl. professional)†	473	491	484
Female youth	2'188	2'204	2'200
<i>Female registered plus youth (sum)</i>	<i>2'661</i>	<i>2'695</i>	<i>2'684</i>
Female not registered	9'008	9'189	9'188
Female children and occasional*	10'000	10'000	10'000
<i>Female not registered (sum)</i>	<i>19'008</i>	<i>19'189</i>	<i>19'188</i>
<i>Female total (sum)*</i>	<i>21'669</i>	<i>21'884</i>	<i>21'872</i>
Total registered (sum)	12'657	13'105	13'083
Total youth (sum)	17'302	17'997	17'772
<i>Total registered plus youth (sum)</i>	<i>29'821</i>	<i>30'975</i>	<i>30'712</i>
Total not registered (sum)	81'460	91'403	98'234
Total children and occasional (sum)*	120'000	120'000	120'000
<i>Total not registered</i>	<i>201'460</i>	<i>211'403</i>	<i>218'234</i>
<i>Total all players (sum)*</i>	<i>231'279</i>	<i>242'378</i>	<i>248'954</i>
Male referees	655	678	673
Female referees	42	42	42
<i>Total referees (sum)*</i>	<i>697</i>	<i>720</i>	<i>715</i>
Male officials	3'479	3'218	3'619
Female officials	440	429	443
<i>Total officials (sum)</i>	<i>3'919</i>	<i>3'647</i>	<i>4'062</i>
<i>Total involved in soccer (sum)</i>	<i>235'895</i>	<i>246'745</i>	<i>253'723</i>

* additional estimate for the number of children and occasional players not included in the associations' estimates (see appendix 1).

† In the *Big Count* professional players were detailed separately. The following numbers were reported or estimated (due to the low numbers involved the values for women could not be estimated):

	Original values	Informed est.	Original plus est.
Men	138	125	141
Women	2	(2)	(2)

Table 2: Different estimates of the number of football clubs and teams based on the data obtained from the *Big Count* figures (Numbers in thousands)

	Original values	Informed estimate	Original values plus estimates
Clubs	280	305	293
Teams	1'467	1'548	1'529

In Table 1 it is also noticeable that the majority of players, both men and women, play the game „Free“, that is to say not as members of a club or association. If we now only consider registered players, that is those who are members of the approximately 300,000 clubs and the 1.5 million teams (c.f. Table 2), then it would appear that around 12.5 million (adult) men and half a million (adult) women play football. In addition there are about 15.5 million male and 2 million female juniors. Thus it can be concluded that about 30 million people play football in an organised environment. While looked at from one angle this is „only“ half a percent of the world’s population, it is, from another point of view, roughly equal to the entire population of Canada.

From Table 1 it can also be seen that while football is still largely a male domain, the proportion of female players has already reached about 9% of the total number involved. Also in the women’s game it seems that football is much more a game for the young than is the case for men: in the latter case the proportion of juniors among those registered is about 55%, while for women it is nearly 82%. This greater percentage among the young can be seen as evidence of the increasing appeal of the game for women.

Table 3: Male and female football players by confederation ("informed estimates")

	UEFA	CONMEBOL	CONCACAF	AFC	CAF	OFC
Male registered	17'116	3'742	3'690	2'240	1'232	260
Male not registered*	32'421	17'615	24'347	96'952	19'907	922
Male total	49'537	21'357	28'037	99'192	21'139	1'182
<i>% of population</i>	<i>6.1</i>	<i>6.3</i>	<i>5.9</i>	<i>2.8</i>	<i>2.7</i>	<i>4.0</i>
Female registered	706	23	1'846	53	34	34
Female not registered*	2'337	667	8'667	6'045	1'411	60
Female total	3'043	690	10'513	6'098	1'445	94
<i>% of population</i>	<i>0.4</i>	<i>0.2</i>	<i>2.2</i>	<i>0.2</i>	<i>0.2</i>	<i>0.3</i>
Total registered	17'822	3'765	5'536	2'293	1'266	294
Total not registered	34'758	18'282	33'014	102'997	21'318	982
Total all players	52'580	22'047	38'550	105'290	22'584	1'276
<i>% of population</i>	<i>6.4</i>	<i>6.5</i>	<i>8.1</i>	<i>3.0</i>	<i>2.9</i>	<i>4.3</i>

* The proportion of "children and occasional players" is based on each confederation’s population as a fraction of the world’s population.

In addition, with respect to women's football and the spread of the game in general, it can be seen from Table 3 that there are considerable regional variations. As the table shows, Europe and America have the most footballers, with CONCACAF (thanks to the USA and Canada) showing high numbers in the women's game. In comparison, the other confederations must be viewed as lagging behind in their development of the women's game. Asia and Africa also show potential for development in terms of the number of men involved. Overall, it seems that CONCACAF - with an estimated 8.1% (one in 12) of the population involved – is the „keenest“ football confederation, while about one person in 15 plays the game at least occasionally in Europe and South America. The figures for AFC, CAF and OFC are well below these levels, lying roughly in the three to four per cent range at the moment.

Conclusions and Outlook

The data presented above shows clearly that football is of great worldwide significance. Around 4% of the total population of the world play football at least occasionally or are involved in the games as referees or in other functions. A further point that can be noted is that there are still considerable regional differences. In view of these it seems that efforts to increase the attraction of the sport should be particularly rewarding in Asia and Africa.

It is also the AFC and CAF areas that show especially low proportions of women involved, which cannot simply be explained – by comparison with Europe and America – on cultural, religious or political conditions. Rather in this connection it can be concluded that on a global level women's football is not attracting enough attention in terms of numbers. However, as the high proportion of younger women compared to the total number of women involved in the game shows, a change seems to be in progress. If it proves possible to keep these females in the sport after they have emerged from childhood or adolescence then this would certainly be a major step in increasing the proportion of women in the game

However, as experience in connection with the *Big Count* has shown, the figures obtained are problematic – despite, or perhaps because of – the major importance of the game. Exactly when the aim is to make predictions about the number of occasional or children players, a lack of reliable data makes it necessary to make a number of additional assumptions. Survey results within Europe have indicated that the „official figures“ submitted by the associations definitely underestimate the actual importance of the game. Therefore it would seem appropriate to collect and evaluate the relevant information more systematically.

In order to obtain definite results that could be published one of the models Table 1 needed to be selected. From a technical point the results given under „informed estimates“ are the most reliable, while the „politically correct“ figures from the third column are at times also close to the original values and of high quality as well.

Appendix 1: Notes about the supplementation and correction of the data

a) Regression estimates

For reasons of clarity linear regression models were used, in which the data values in question (e.g. the number of registered players) were estimated as a function of a number of so-called independent variable. The model had the following general form:

$$y=b + a_1*z_1 + a_2 * z_2 + \dots + a_x * z_x$$

where: y: dependent variable (e.g. number of registered players)

b: constant

a₁,a₂,..., a_x: regression weightings

z₁,z₂,..., z_x: independent variables (e.g. population)

The following independent variables were used:

- Total population (from the World Development Report of the World Bank 2000, partly supplemented by data from the World Fact Book).

Assumption: As population increases so does the number of players (male and female)

- Gross Domestic Product per capita (adjusted to reflect purchasing power parity, based on the World Development Report of the World Bank 2000, partially supplement by figures from the World Fact Book).

Assumption: With increasing wealth, the possibility of playing football in an organised environment increases.

- Membership of a particular confederation.

Assumption: Football has a different status in different confederations. This must be taken into account in the estimates. For example, it can be assumed that the probability that a person plays football in the CONMEBOL region is higher than within AFC. The different confederations were entered as so-called dummy variables in the regression equation, with the values 0 (= association does not belong to the appropriate confederation) and 1 (= association does belong to its confederation). Overall five variables - UEFA, CONMEBOL, CONCACAF, AFC and CAF - were used. The value for OFC was set at 0 for statistical reasons (if an additional variable for OFC had been entered then the model would have been overdetermined).

- Date of foundation of the association (according to FIFA data).

Assumption: the older an association the more firmly football is anchored in that society and therefore the greater the number of players (m and f).

The first regression estimates obtained using this basic model brought to light three problems that had to be solved:

- 1st Problem: The relationships between the dependent and the various independent variables (population, Gross Domestic Product per capita) are not linear in the sense of the regression model.

Solution: Via a logarithmic transformation the doubtful variables were changed into an approximately linear form. In the regression model therefore, the logarithmic values of the dependent variables were estimated on the basis of logarithmic (populations, GDP per capita) and non-logarithmic figures (confederations, date of foundation).

2nd Problem: The fact that the original data was not plausible in many cases led to the regression estimates for the so-called „outliers‘ being distorted. In the example of Argentina mentioned above, the figure lies well below the regression line and thus distorts all the other estimates.

Solution: To overcome this the regression model was calculated in two stages: in a first step an estimate was made on the basis of all the available data. Figures that deviated significantly from this estimate (outliers) were eliminated for the second estimate. In other words, for the new estimate only the „average“ associations were considered. The resultant model then served as a basis for the estimation of missing and implausible data.

3rd Problem: For many associations there were no responses about GDP per capita. This is due on the one hand to the fact that in many cases the figures are not available, and on the other that the corresponding figures are known at national level but often cannot be reliably determined for associations that are just a part of the larger nation. The regression estimates including GDP and later estimates with data missing are therefore incomplete, due to this lack of information. However, it was observed that the GDP per capita is in fact an important predictor.

Solution: In view of these problems, two estimates were calculated, one with and one without an indication of GDP per capita. Thus for many countries there are two figures for missing and implausible data, which were also used in the subsequent corrections.

In brief, for each dependent variable the following procedure was chosen:

- a) Calculation of two regression estimates in each case (with and without GDP per capita) for the questionable dependent variables.
- b) Elimination of outliers and re-calculation of the regression estimates (c.f. Table A1)
- c) Insertion of the independent variables in the two regression estimates und calculation of the estimates, then conversion back out of logarithmic form.
- d) Checking the estimates or comparing them with the original values; insertion of missing data and correction of given values, if needed, on the basis of the estimates and comparison with similar associations.

Table A1 presents first an overview of the various regression estimates, from which can be seen that most of the models re-produce the logarithmic original values very well, with a degree of explained variance between 80 and over 90 per cent. However, it is noticeable that this level of explained variance has to be obtained in the definite models with a relatively small number of cases. This indicates that in each case a relatively large number of outliers exists, which have to be checked exactly when running a control.

In addition the table shows that the estimates for women are significantly less accurate, and are based on even fewer cases than for the men. This is because in a number of confederations there are few, sometimes no, women playing football, so that naturally less original information is available. Where such information is available, it varies more than it does for men; the range is from real „women’s“ football associations like the USA or Canada, down to associations – especially from AFC – where women’s football plays no role at all or only a very marginal one.

Overall, however, it can be considered that the regression estimates are good, and that they provide a reliable basis for the correction of data and for the estimation of missing figures.

Table A1: Figures for the regression estimates (models, in which the "outliers" have been removed)

	Regression estimate with GDP per capita		Regression estimate without GDP per capita	
	No. of cases	R ²	No. of cases	R ²
Male professional	78	0.84	76	0.83
Male registered	92	0.86	107	0.86
Male youth	104	0.88	114	0.85
Male registered plus youth	98	0.94	112	0.90
Male not registered	79	0.91	96	0.89
Male total	100	0.92	115	0.92
Female professional	*	*	*	*
Female registered	75	0.84	83	0.67
Female youth	58	0.81	72	0.69
Female registered plus youth	81	0.81	90	0.70
Female not registered	50	0.72	61	0.68
Female total	89	0.74	97	0.74
Male referees	91	0.88	108	0.86
Female referees	65	0.64	76	0.59
Male officials	81	0.82	95	0.79
Female officials	55	0.59	67	0.57
Clubs	81	0.85	92	0.88
Teams	97	0.92	110	0.91

*Estimates not possible - not enough cases

Supplementary estimates for non-registered players

Many associations had trouble in estimating the number of non-registered players, so that it seemed necessary to supplement the associations' responses. As the overview of the data shows, most of the associations seem to have implicitly assumed that what was wanted involved a more or less regular activity for non-registered players, and children seem to have been systematically forgotten.

Only a very few associations allowed themselves estimates that seemed on the high side, which probably did contain children and occasional players. On the male side only Mali (46 %), Paraguay (28 %), Ecuador (16 %), Mexico (13 %), Cuba (9 %), Bangladesh (8 %), Zambia (7 %), Lebanon (7 %) und Italy (5 %) reported a proportion of non-registered players of five per cent or more of the population. On average, however, the proportion of non-registered players proved to be just over a half of one per cent of the population. And in the women's game, the associations' estimates were above two per cent of the population in only four cases (USA, Mexico, Eritrea, Turks and Caicos Islands) – the average of the responses here was only 0.03 %!

However, it is reasonable to assume that the number of people that actually play football from time to time is higher. In addition the number of children who play outside of club sport has been clearly under-estimated by the associations.

In order to get a better estimate of occasional footballers and children, a number of additional sources were contacted in January, and also, via E-mail a questionnaire from about 20 international and sports experts was started. Here different answers are being collected and experience gathered which would indicate that the responses to the *Big Count* were indeed too low.

As one example, a comparison can be made with figures for Europe from the COMPASS-Project, which showed that in Great Britain around 5 million people play football on at least an occasional basis, a value about 40 % higher than the figure provided by the associations and our additional estimates. A similar result was obtained in the case of Switzerland, in which nearly 800,000 people reported having played football at least once in the previous year. The corresponding figure from the *Big Count* is however only approximately 440'000. Further comparisons with data from Finland, Germany and South Korea present a similar picture; the results obtained from the surveys always indicate a considerably higher total.

Germany and Finland are also the only countries for which data about children and young people playing football are available. In Germany some 30 % of the 14-18 age group say that they play football (this represents 56 % of the boys active in sport and 9 % of the girls who take part in sports). And for Finland the figures indicate a participation of 18 % of the 3-18 age group involved in football (31 % of the boys and 5 % of the girls). Once again these figures are quite a lot higher than those submitted by the associations under „children involved in football“.

These comparisons would indicate that some adjustment of the response data is called for, but precisely because the comparisons are available in only a limited number of cases, it is not possible to make reliable estimates. Therefore it makes sense here to work with „scenario estimates“, two of which are presented in Table A2.

Scenario 1 is based on a conservative assumption that football is to a large extent a men’s sport and that the game is less widely played in other regions of the world than in Europe, for which the additional data is available. Scenario 2 supposes that playing the occasional game of football is a quite common event (for example, at least once a year with friends on a Sunday afternoon) and that women are also likely to be involved.

Table A2: Scenarios for estimating the numbers of children and occasional adults involved in football

	Scenario 1 Assumption	Estimate (in Mio.)	Scenario 2 Assumption	Estimate (in Mio.)
Children male	10 % of boys	90	25 % of boys	225
Occasional male	1 % adult males	18	2 % of adult males	36
<i>Total male</i>		<i>108</i>		<i>261</i>
Children female	1 % of all girls	9	5 % of all girls	45
Occasional female	0.05 % of adult females	1	0.5 % of adult females	9
<i>Total female</i>		<i>10</i>		<i>54</i>
<i>Total children and occasional</i>		<i>118</i>		<i>315</i>

Basis for the calculation: a total world population of 6 billion people and a proportion of 30% of them under 15 years of age (according to UN Population Statistics).

In order not to run the risk of providing estimates that might be seen as over-optimistic, it seems more appropriate to work with the more conservative Scenario 1, and the results obtained in this manner have been used previously in Table 1. Scenario 2 however has definite heuristic value and would indicate that the total number of people who at least once a year could be found kicking a ball is around 200 million higher. In other words, the total presented in Table 1 of 250 million could in fact be almost twice as big under different assumptions. In order to get closer to the true figures, further research would be necessary.