Delpeyrou, Léonie; Bertoli, Simone

Working Paper

"Building Nations through Shared Experiences: Evidence from African Football": a Replication of Depetris-Chauvin, Durante and Campante (2020)

I4R Discussion Paper Series, No. 193

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Delpeyrou, Léonie; Bertoli, Simone (2024) : "Building Nations through Shared Experiences: Evidence from African Football": a Replication of Depetris-Chauvin, Durante and Campante (2020), I4R Discussion Paper Series, No. 193, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/307961

INSTITUTE for REPLICATION

No. 193
DISCUSSION PAPER SERIES

"Building Nations through Shared Experiences: Evidence from African Football": a Replication of Depetris-Chauvin, Durante and Campante (2020)

Léonie Delpeyrou

Simone Bertoli

This paper has received two responses:

Campante, Filipe, Emilio Depetris-Chauvin, and Ruben Durante. 2024. "Building Nations through Shared Experiences: Evidence from African Football": A Reply to Delpeyrou and Bertoli (2024). 2024. *I4R Discussion Paper Series No. 194*. Institute for Replication.

and: https://osf.io/gj7v3/

December 2024

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 193

"Building Nations through Shared Experiences: Evidence from African Football": a replication of Depetris-Chauvin, Durante and Campante (2020)

Léonie Delpeyrou[1], Simone Bertoli[2,3,4]

[1]Université de Lausanne/Switzerland
[2]Université Clermont Auvergne, Clermont/France
[3]Centre d'Études et de Recherches sur les Développement International (CERDI), Clermont/France
[4]Centre National de la Recherche Scientifique (CNRS), Paris/France
[5]Institut de Recherche pour le Développement, Marseille/France

DECEMBER 2024

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

"Building Nations through Shared Experiences: Evidence from African Football": a replication of Depetris-Chauvin, Durante and Campante (2020)*

Léonie Delpeyrou[a] and Simone Bertoli[b]

[a] *Université de Lausanne, IDHEAP, CH-1015 Lausanne*

[b] *Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000, Clermont-Ferrand*

Abstract

Depetris-Chauvin et al. (2020) analyze the effects of football on civil conflict in Africa. We engage into a replication of their analysis. This unveils major inconsistencies between the sample selection criteria described in the article and the composition of the estimation sample (41 errors for 109 country × qualification campaign pairs). Once these inconsistencies are removed, the evidence on the effect of a close qualification to the African Cup of Nations on conflict becomes substantially weaker. Several specifications produce a non-significant coefficient for the treatment. The placebo test proposed by the Authors themselves casts doubts on the validity of the identifying assumption in most specifications with a significant effect of the treatment.

---

*Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante, "Building Nations through Shared Experiences: Evidence from African Football." *American Economic Review*, 2020, 110 (5): 1572-1602; the data and the code underlying this comment are available at: https://www.dropbox.com/scl/fo/lk7akxuy82rpews8b8jq2/AFuSJm69qeI1Jq585_qwm64?rlkey=j0tjju8iu1botoslymr9m1jbx&st=4adnj4v4&dl=0; the Authors are grateful to Filipe Campante and Emilio Depetris-Chauvin and Ruben Durante for their reply to an earlier version of our paper, and to Clément de Chaisemartin, Vianney Dequiedt, Jesús Fernández-Huertas Moraga, Jules Gazeaud, David McKenzie and Èric Roca Fernández for their comments; Simone Bertoli acknowledges the support received from the French government through the investment program "France 2030" (ANR-16-IDEX-0001); the order of the Authors has been determined using the AEA Author Randomization Tool (confirmation code Kl pZtQ2abHEu); the usual disclaimers apply.

Simone Bertoli, UCA, CNRS, IRD, CERDI, 26 Av. Léon Blum, F-63000, Clermont-Ferrand; email: simone.bertoli@uca.fr (corresponding author).

1 Introduction

Depetris-Chauvin, Durante and Campante (2020), DC^2 henceforth, provide an econometric analysis of the role of (male) football in fostering a shared national identity over distinct ethnic identities in Sub-Saharan African countries, and of the ensuing reduction in the incidence of civil conflicts. The second part of their econometric analysis explores whether countries that obtained a close qualification for the final phase of the Africa Cup of Nations (ACN) experience, in the following six months, a reduction in the intensity of civil conflict, compared to countries that were in a position to qualify before the last match day, but were eventually eliminated. The analysis draws on 10 editions of the qualifications group stage of the ACN, from 1998 to 2015, with the estimation sample being composed by the 109 country-qualification campaign pairs that "(barely) qualified" or "(barely) did not", as DC^2 describe them, out of a total of 322 country-qualification campaign pairs. DC^2 write that "[t]he underlying identification assumption is that if two teams in the same group got to the last match day with concrete chances of qualifying, which one would actually qualify will be determined by *quasi-random circumstances*, such as a *goal scored in the final minutes of the last match by one side or the other*." (p. 1591, emphasis added). The estimates reveal a significant reduction in the number of conflict events, and in particular of ethnic-related conflict. Google Scholar records 302 citations of DC^2 (information retrieved on November 25, 2024).

We engage into a direct and conceptual replication of the country-level analysis in DC^2. Our replication exercise is confronted with a major analytical challenge arising from three closely related factors. First, the code and the data used to define the treatment and the control group do not belong to the replication package, which only includes an Excel file with the list of the country-qualification campaign pairs in the treatment and in the control group, and the treatment date associated to each of them. Second, no formal definition of the sample selection criteria is provided, neither in the main text nor in the online Appendix. Third, the brief descriptions provided of the sample selection criteria in the original article are contradictory, and inconsistent with the actual composition of their estimation sample.

As an initial step, we carefully analyze the four brief descriptions of the sample selection criteria provided by DC^2, using the data to clarify the instances in which these definitions are contradictory (see Section A in the online Appendix). Then, we undertake a substantial effort of data collection and coding, which rests on an explicit theoretical definition of the

sample selection criteria, to define the estimation sample in a consistent and theoretically sound way. First, we gather information about the format of each qualification group stage of the ACN (see Section B), including the set of tiebreaking criteria used for each edition, which are retrieved using the Wayback Machine of the Internet Archive (see https://web.archive.org/) applied to the website of the CAF, the African Football Confederation (see Section C). Second, we create a dataset including the table of each group before the last match day, the matches to be played on the last match day, and their dates. Third, we present a formal definition of the treatment and control groups, and of the treatment date, which encompasses the specificities of each edition (see Section D). Fourth, this formal definition is implemented in a code that uses the data that we have collected to automatically derive the composition of the estimation sample.

We include in the treatment (control) group country-qualification campaign pairs that obtained a qualification (failed to qualify) for the final phase of the ACN if and only if there was at least a combination of results in all groups on the last match day that would have eliminated (qualified) them. These sample selection criteria that we implement in our code do not differ from DC^2: they are simply based on a thorough analysis of the descriptions provided in the original article. Thus, our code allows to deploy the same identification strategy introduced by DC^2 on a correctly defined estimation sample. This analysis reveals five inclusion errors in the original estimation sample, corresponding to cases in which the qualification status of a country was fully determined before the last match day. We also uncover 36 country-qualification campaign pairs with an undetermined qualification status before the last match day which do *not* belong to the original estimation sample. We also document that the original estimation sample included several country-qualification campaign pairs with a qualification status which was virtually fully determined before the last match day. To give a telling example, we can observe that DC^2 include Madagascar 2004 in the control group, even though this country needed to win by a margin of at least 17 goals on the last match day to qualify (see Section A.2.2). This margin is seven goals above the largest margin of victory ever observed in ACN qualifiers, corresponding to São Tomé-et-Principe vs. Nigeria 0-10, played on June 13, 2022 (Jürisoo, 2024).

This, in turn, reveals that when DC^2 write that only country-qualification campaign pairs with "concrete chances of qualifying" (p. 1591) are included in the estimation sample, this can only be interpreted as referring to teams with nonzero chances of qualification (or

elimination) before the last match day.[1] Therefore, we refer to the 36 country-qualification campaign pairs with an undetermined qualification status before the last match day not included in the original estimation sample as exclusion errors.

Furthermore, our analysis of the data reveals multiple independent instances in which DC[2] treat differently country-qualification campaign pairs that were exactly in the same situation before the last match day, an inconsistency that our code clearly avoids. In particular, the original estimation sample includes a "twin" for six country-qualification pairs that represent exclusion errors. A twin is a team belonging to the estimation sample in DC[2] which was included in the same group, and which had the same number of points before the last match day. For instance, DC[2] include Rwanda 2004 in the estimation sample, while they exclude Ghana 2004, albeit both teams where included in Group 13 and had 4 points before the last match day, which featured a direct confrontation between the two (twin) teams, which determined the qualification for the final phase of the ACN (see Section G.17).[2] Moreover, the original estimation sample also excluded several country-qualification campaign pairs, which do not have a twin, that unambiguously had substantial chances of qualification or elimination before the last match day. For instance, DC[2] exclude Kenya 2002 from the control group, even though Kenya would have certainly qualified if it had not lost on the last match day (see Section I.4). Similarly, DC[2] also exclude Mozambique 2000 from the estimation sample (see Section I.1), even though its elimination was determined by a goal scored by Eritrea 9 minutes before the end of its last match, a situation that seems to fit perfectly with the "quasi-random circumstances, such as a goal scored in the final minutes of the last match by one side or the other" that should determined the qualification status of a team included in the estimation sample (DC[2], p. 1591).[3]

[1] It is conceptually extremely difficult to define a measure of the chances of qualification of a team before the last match day, as this would require defining the probability distribution of the results of all the matches that contribute to determine the qualification status of a team; it would have been, therefore, surprising if DC[2] had used such a measure, which is also very demanding in terms of data, in the definition of the estimation sample without defining or even describing it in the original article.

[2] Interestingly, Ghana would have qualified by winning or drawing against Rwanda, while Rwanda could qualify only by winning against Ghana, as it did; thus, the set of results that could qualify Rwanda was actually smaller than the one qualifying Ghana.

[3] Section I provides seven examples of country-qualification campaign pairs, which do not have a twin in the original estimation sample, and that are not included in the analysis even though they had concrete chances of qualification before the last match day.

DC^2 present, in Tables 5 to 8, 20 different variants of Eq. (2)[4], which is used in the country-level analysis, and Table 5 also presents a placebo test on the parallel trend assumption underpinning their difference-in-difference specification. We use the estimation sample of 140 country-qualification campaign pairs that we have derived from the application of a formal definition of the sample selection criteria to re-estimate all the specifications of Eq. (2) in the main text of DC^2, and to verify, for each specification, the validity of the identifying assumption through the placebo test that DC^2 have proposed. Using the correct sample and treatment dates and the same specification of Eq. (2) as in DC^2,[5] the empirical evidence of a negative effect of a close qualification to the ACN on conflict intensity is weaker, and the parallel trend assumption is often violated. Notably, only four out of the 20 specifications (*i*) the coefficient of the treatment is negative and significant, and (*ii*) the coefficient of the placebo treatment is non-significant. 11 additional specifications produce a negative and significant coefficient of the treatment, but for seven of them the placebo treatment is significant, thus calling into question the validity of the identifying assumption, and for four additional specifications it is only marginally non-significant (p-value between 0.111 and 0.145). Conversely, all the five specifications of Eq. (2) for which the treatment is not significant are associated with a satisfactory placebo test.

A corrigendum of the original article was published by the *AER* on December 3, 2024 (Depetris-Chauvin et al., 2024). This corrigendum is based, as the Authors acknowledge, on an earlier version of our paper, which was submitted to the *AER* in July 2024, and shared with the Authors one month before the submission. This corrigendum acknowledges several of the points that we had raised, and it provides some additional details on the sample selection criteria that were missing in DC^2. This corrigendum does *not* address the key

[4]These variants relate to several dimensions, such as the inclusion of four lags of the dependent variable, the estimator that is used, the estimation on a subset of observations, to the interaction of the treatment with another variable, or to different definitions of the dependent variable, which only consider conflict events with specific features.

[5]We also correct three independent mistakes in the code included in the replication package, but this still means that we rely on the same specification as in the original article, according to Clemens (2017): "[t]he 'same' specification, population, or sample means the same as reported in the original paper, not necessarily what was contained in the code and data used by the original paper. Thus for example if code used in the original paper contains an error such that it does not run exactly the regressions that the original paper said it does, new code that fixes the error is nevertheless using the 'same' specifications (as described in the paper)." (p. 328).

issues that we had raised, and it presents the estimates for the country-level analysis that are obtained from a sample that is inconsistent both with the description of the estimation sample provided in DC[2], and with the sample selection criteria detailed in the corrigendum itself.

The rest of the paper is composed by five short sections: Section 2 provides a formal definition of the sample selection criteria, and of the treatment date. Section 3 applies these definitions to obtain a theoretically correct and consistent composition of the estimation sample. Section 4 uses this correct sample to replicate the country-level analysis, testing also the validity of the underlying identifying assumption for each specification. Section 5 describes the content of the corrigendum of the original article. Then, Section 6 draws the main conclusions of our replication exercise.

2 A formal definition

Our initial analysis of the descriptions of the estimation sample provided by DC[2] and of its actual composition (see Section A) has clearly revealed that this should include country-qualification pairs with an undetermined qualification status before the last match day. This, in turn, informs our formal definition of the sample selection criteria and of the treatment dates.

2.1 Sample selection criteria

Section D in the online Appendix provides a formal mathematical definition of the sample selection criteria, which encompasses the great variety of formats of the various editions of the qualification group stage for the ACN. This definition is particularly challenging for the four editions (2004, 2008, 2012 and 2015) including a virtual playoff,[6] as it requires to analyze the final tables in the other groups corresponding to any possible combination of results on the last match day. For instance, for the qualification for the ACN 2012, if we just consider the number of points obtained by the home team in each match, this gives us $3^{22} \approx 31.4$ billion combinations, as 22 matches were played in total on the last match day in 11 groups where the three best runner-ups obtained a qualification for the final phase.

[6] A virtual playoff simply gathers, say, the runner-ups of various groups, qualifying one or more of these teams to the final phase of the ACN, on the basis of their ranking in this virtual group.

The definition of the sample selection criteria uses the table before the last match day, the program of the matches on the last match day, the information about the format, including the set of tiebreaking criteria, for each edition of the qualification group stage. A team that eventually qualified (failed to qualify) for the final phase of the ACN is included in the estimation sample if and only if there is at least a combination of results on the last match day such that the team would have ended up in a position in the final table of its own group which (*i*) implied a direct elimination (qualification), (*ii*) gave access to a virtual playoff group with a number of points that were insufficient to guarantee an indirect qualification (sufficient to guarantee an indirect qualification) via the virtual playoff group, given all possible combination of results in the other groups, or (*iii*) gave access to a real playoff group.[7] The qualification status is initially assessed under the analytically convenient assumptions that ties between two or more teams in the final table are broken with a random draw. The assumption that ties are broken with a random draw is immaterial for a team for which there is at least a combination of results satisfying one of these three conditions that does not involve a tie in the final table. If the only combination(s) of results on the last match day satisfying one of these three conditions involve(s) a tie, we verify that this condition can be satisfied (given the results up to the last match day) even when applying the actual set of tiebreaking criteria for each edition. This last step is crucial in several instances (see, for instance, South Africa 2006 in Section F.2).

2.2 Treatment dates

Section H in the online Appendix describes in detail how we define the treatment dates. This involves conceptual challenges for two editions (2000 and 2008), whose specific formats dive a wedge between the treatment date and the date at which we analyze whether the qualification status of a team was undetermined. In 2000 edition, the qualification status of the three teams joining a real playoff group was determined after six additional matches played in the two months following the last match day of the group stage, while in the 2008 edition the qualification status of some of the teams in the virtual playoff group was determined only after the last matches of Groups 2 and 9, which intervened five weeks after the last matches of the other groups.

[7] Any team joining a real playoff group could be either eliminated or qualified.

3 Correct sample and treatment dates

DC^2 identify 109 country-qualification campaign pairs in the estimation sample, out of which 49 are treated and 60 are in the control group. Once we apply our formal definition to the raw data that we have gathered, we uncover many errors in the composition of the estimation sample: 41 errors for 109 country-qualification campaign pairs.

Notably, DC^2 incorrectly include in the estimation sample five country-qualification campaign pairs (see Sections F.1-F.5) for which the qualification status was fully determined before the last match day. Thus, inclusion errors represent $5/84 = 6.0$ percent of the original estimation sample corresponding to editions of the ACN featuring a group qualification stage.[8] Furthermore, DC^2 exclude from the estimation sample 36 country-qualification campaign pairs (see Sections G.1-G.36) whose qualification status was still undetermined.[9,10] Exclusion errors represent $36/156 = 23.1$ percent of the country-qualification campaign pairs not included in the original estimation sample.[11] The distribution of these errors across editions of the ACN and between the treated and control group is provided in Table 1, where a R and a V denote the editions with a real and a virtual playoff group respectively.[12]

The only edition without any error is the ACN 2013, which is unique as it did not feature a group stage (see Section B.9), so that the definition of the treatment and control group for this edition is straightforward. 33 out of the 41 errors in Table 1 relate to the five qualification campaigns including either a real or a virtual playoff group; more precisely, 28 mistakes relate to the four editions of the ACN with a virtual playoff, as this substantially complicates the identification of the teams with an undetermined qualification status before the last match day, requiring a comparison across groups. For instance, DC^2 included Sudan

[8] All editions except the ACN 2013 (see Section B.9), for which all 25 Sub-Saharan countries playing in two-legged direct elimination matches.

[9] As discussed in the Introduction, we refer to these cases as exclusion errors, as their classification is inconsistent with the verbal descriptions of the estimation sample provided by DC^2.

[10] Two additional exclusion errors relate to the qualification for the World Cup, which DC^2 use in Table A.28 (see Section K).

[11] There are 265 country-qualification campaign pairs corresponding to Sub-Saharan African countries for which ACLED data are available, and are thus potentially included in the estimation sample; 109 are actually included by DC^2 in the estimation sample.

[12] A virtual playoff group simply compares the points obtained by the various teams in their own groups, possibly adjusted (in the 2012 edition) to account for the different number of matches played in the various groups.

Table 1: Mistakes in the definition of the sample for Eq. (2)

	Treated			Control			Estimation sample			
		Errors			Errors			Errors		
Edition	DC^2	Incl.	Excl.	DC^2	Incl.	Excl.	DC^2	Incl.	Excl.	Total
1998	4	0	1	6	0	1	10	0	2	2
2000^R	3	0	1	4	0	4	7	0	5	5
2002	3	0	0	5	0	2	8	0	2	2
2004^V	7	0	2	6	1	6	13	1	8	9
2006	2	1	0	1	0	0	3	1	0	1
2008^V	7	1	2	7	0	6	14	1	8	9
2010	2	0	2	2	0	1	4	0	3	3
2012^V	5	0	2	10	1	2	15	1	4	5
2013	11	0	0	14	0	0	25	0	0	0
2015^V	5	1	2	5	0	2	10	1	3	5
Total	49	3	12	60	2	24	109	5	36	41
		15			26			41		

in the treatment group for the ACN 2008, but Sudan was certain to qualify either as the winner of its group or as one of the best three runner-ups via the virtual playoff group, irrespective of the results of the last match day in all other groups (see Section F.3). The five exclusion errors for the ACN 2000 all relate to countries in groups giving access to the real playoff group introduced by the CAF after the change of the host country for the final phase (see Section B.2), a change in the format that DC^2 seem to have not taken into account. The 41 errors in Table 1 relate to both the treatment (15 errors) and the control group (26 errors).

An analysis of the exclusion errors also reveals that our code meets a natural requirement for a sample selection criterion, that DC^2 fail to achieve: a consistent classification of country-qualification campaign pairs which were in an identical situation before the last match day. Notably, DC^2 treat inconsistently six couples (or triplets) of country-qualification campaign pairs, which we refer to as "twins". Twins were included in the same group, and had the same number of points before the last match day, but DC^2 excluded from the estimation sample one of the twins, while including the other(s). The six couples or triplets are: Tanzania 1998 (excluded) and Liberia 1998 (control) in Group 6 (see Section G.2), Namibia 2002 (excluded), Zambia 2002 (treated) and Madagascar 2002 (control) in Group 1 (see Section

G.8), Ghana 2004 (excluded) and Rwanda 2004 (treated) in Group 13 (see Section G.17), Tanzania 2008 (excluded) and Senegal 2008 (treated) in Group 7 (see Section G.20), Tanzania 2012 (excluded) and Algeria 2012 (control) in Group D (see Section G.30) and Ghana 2012 (excluded) and Sudan 2012 (treated) in Group I (see Section G.32).[13] In three of these cases, related to Tanzania 1998, Ghana 2004 and 2012, the last match day even featured a direct confrontation between the twins. Similarly, DC^2 inconsistently treat Angola 2015 and Mozambique 2015, which had both 5 points before the last match day, and they might have only qualified via the virtual playoff group by winning the last match in their respective groups.[14] Our code treats consistently all these cases, and it also avoids the inconsistent treatment of the teams not playing on the last match day that we described in Section A.1. Furthermore, we also uncover that DC^2 not only include in the estimation sample country-qualification campaign pairs with a virtually determined qualification status (see Section A.2), but they also exclude many instances in which a country unambiguously had substantial chances of qualifying (see Section I). For instance, DC^2 exclude Ethiopia 2004, from the control group, event though Ethiopia would have qualified if it had won or drawn its last match.

Out of the 104 country-qualification campaign pairs correctly included in the estimation sample by DC^2, the treatment date is wrong by five weeks in two cases. Furthermore, five country-qualification campaign pairs have a treatment date that is five to nine weeks away from the one of the other countries included in the estimation sample for these two qualification campaigns (see Section H).

4 Estimates and placebo tests

DC^2 draw, for each of the 109 country-qualification campaign pairs (c, q) in the estimation sample, on the data on conflict events from the Armed Conflict Location and Event Data Project (ACLED) dataset for a period of 25 weeks before and after the treatment date to

[13]The inclusion of the twins in the estimation sample can be verified from Table A.20 in the online Appendix of DC^2, and in the file `classification.xls` in the replication package of DC^2 for Algeria 2012.

[14]See G.34 for Angola, and https://athlet.org/football/can/2015/qualifiers/group-f for Mozambique, which is included by DC^2 in the control group.

estimate the following equation:

$$\text{Conf}_{c,q,t} = \alpha + \beta \text{Post\_Qual}_{c,q,t} + \sum_{k=1}^{4} \delta^k \text{Conf}_{c,q,t-k} + \sum_{t=-25}^{25} \Gamma_t + \Pi_{m_{c,q}(t)} + \Delta_{c,q} + \epsilon_{c,q,t} \quad (2)$$

where c, q, and t denote respectively country, qualification campaign, and time to and since the qualification (or elimination). The dependent variable $\text{Conf}_{c,q,t}$ is a measure of conflict intensity, and $\text{Post\_Qual}_{c,q,t}$ is a dummy equal to 1 in the 25 weeks following a *close* qualification, and 0 otherwise. Γ_t, $\Pi_{m_{c,q}(t)}$ and $\Delta_{c,q}$ are respectively a set of weekly dummies, calendar-month, and country $\times$ qualification campaign fixed effects.

DC^2 also estimate a placebo, to validate the parallel trend assumption that underpins the identification of Eq. (2). This is described by the following equation:

$$\text{Conf}_{c,q,t} = \alpha + \beta \text{Placebo}_{c,q,t} + \sum_{k=1}^{4} \delta^k \text{Conf}_{c,q,t-k} + \sum_{t=-25}^{-1} \Gamma_t + \Pi_{m_{c,q}(t)} + \Delta_{c,q} + \epsilon_{c,q,t} \quad (3)$$

which is estimated for the 25 weeks before the treatment date for all the country-qualification campaign pairs in the estimation sample, and where $\text{Placebo}_{c,q,t}$ is equal to 1 for treated countries in the 12 weeks before the treatment date, and 0 otherwise.

We re-estimate all the 20 different specifications of Eq. (2) that are presented in the main text by DC^2 (Tables 5-8) using the correct estimation samples and treatment dates[15], and also correcting three distinct mistakes in the code (see Section J). For each of these specifications, we also estimate the placebo test proposed by DC^2 themselves, but that was applied only to the benchmark specification in the original article.

The implications of the incorrect sample and treatment dates on the estimates of Eq. (2) are, *a priori*, ambiguous. The estimation sample used by DC^2 contains 31 fewer country-qualification campaign pairs than the correct sample. This, coupled with the mistakes in a few treatment dates, might reduce the precision of the estimates of Eq. (2). However, if the country-qualification campaign pairs that have been incorrectly excluded or included in the estimation sample by DC^2 happen to be non-randomly selected with respect to unobserved determinants of the evolution of civil conflicts over the year around the treatment date, then the estimates of Eq. (2) might be biased. As DC^2 estimate Eq. (2) on various subsamples, or interact the post-qualification dummy with other variables, the relative importance of the loss of precision and the possible bias might vary across the different specifications.

[15] Tables 5 to 8 in DC^2 include a total of 23 data columns, but Cols. (3) and (4) in Table 5 report the results of the placebo test, and Col. (1) in Table 6 simply reproduces Col. (2) in Table 5.

Tables L.1-L.4 in the online Appendix have the same structure as the corresponding tables and figure in DC^2, except that they also report the estimated coefficient, standard error and p-value of the placebo treatment in Eq. (3) corresponding to each specification. The use of the correct sample and treatment dates results in a weaker empirical evidence about the conflict-reducing effect of a close qualification to the ACN, with a significant coefficient for the placebo treatment in many specifications, calling into question the validity of the parallel trend assumption that underpins Eq. (2). For instance, the estimated coefficient for the treatment stands at -0.132 (s.e. 0.059), with an associated p-value of 0.027 in the first data column of Table L.1, and the coefficient of the placebo treatment for this specification stands at 0.107 (s.e. 0.058), with a p-value of 0.065.[16]

Only four out of the 20 specifications produce a negative and significant coefficient for the treatment, combined with a satisfactory result of the placebo test, i.e., a non-significant coefficient for the placebo treatment. This occurs for Col. (8) in Table L.1, Col. (2) in Table L.2 and Col. (3) and (4) in L.4, with three out of these four coefficients being significant at the 10 percent confidence level. The estimated coefficient for the treatment is negative and significant in 11 additional specifications in Tables L.1-L.4, but the corresponding placebo treatment is also significant for seven of these specifications, and just marginally non-significant in four cases (p-value between 0.111 and 0.145). Interestingly, the placebo is always satisfactory, i.e., not significant, for all the five specifications in which the treatment is not significant. We also re-estimated the placebo test for all these 20 specifications using the original estimation sample of 109 country-qualification pairs, and correcting the three mistakes described in Section J; the coefficient of the placebo treatment is not significant (at the 10 percent confidence level) in 19 out of 20 cases, thus revealing that it is the use of the correct estimation sample and treatment dates that matters in our results.

DC^2 describe as follows the rationale of the placebo test, and of the results of the placebo for the benchmark specification:

> "If conflict was evolving differently in the two groups in the pre-qualification period, which would threaten the validity of our difference-in-differences approach, we would expect the fictitious treatment to display a significant coefficient. The results, in columns 3 and 4 of Table 5, point against this possibility. In particular, *there is no indication that conflict was decreasing*

[16] Using the original code and estimation sample, the coefficient of the placebo treatment for this specification stands at 0.046 (s.e. 0.062), with a p-value of 0.457 (see the third data column in Table 5 in DC^2).

> *in countries that would eventually qualify, relative to countries that would not.*" (DC^2, p. 1592, emphasis added).

The concern expressed by DC^2 about a possible violation of the identifying assumption seems to be centered around the possibility of a reverse causality from a higher conflict intensity to the treatment: countries with deteriorating security conditions might have lower chances of obtaining a (close) qualification to the ACN,[17] something that could result in a negative coefficient for the placebo treatment. However, also a positive and significant coefficient for the placebo treatment constitutes a violation of the identifying assumption. Indeed, in all instances in which the placebo treatment is significant, its coefficient is positive, revealing that treated countries had a significantly higher conflict intensity than control countries before the treatment date.[18] This, in turn, determines an overestimation of the negative effect of the treatment on conflict intensity in Eq. (2).

5 About the corrigendum

The *AER* published an online corrigendum to the original article on December 3, 2024 (Depetris-Chauvin et al., 2024). As far as the country-level analysis in DC^2 is concerned, the corrigendum acknowledges that "in describing the empirical strategy based on close qualification to the ACN finals, we did not provide a sufficiently clear description of how we selected the sample of close qualification cases" (p. 2), and it describes the algorithm that they had applied in DC^2. The corrigendum also acknowledges and corrects two of the three independent coding mistakes described in Section J,[19] and it also acknowledges the mistakes related to the treatment dates for three of the country-qualification campaign pairs in Section H. Furthermore, the corrigendum refers to "idiosyncratic quirks of the qualification process" (p. 2) or to "idiosyncrasies of the tie-breaking criteria for qualification" (p. 3) to explain why a correct definition of the estimation sample was challenging, and to explain

[17] Indeed, the regulations of the ACN include articles dedicated to countries in a state of war and, more generally, referring to "internal situations in a country that may affect security conditions when holding a match"; the Organizing Committee of the CAF can decide, in cases of concerns about security, to impose to the home team that a match of the qualification group stage is organized in another country.

[18] This pattern is also revealed by Figure L.1 in the online Appendix.

[19] The estimates presented in the corrigendum still unnecessarily drop initial observations when including four lags of the dependent variable.

the instances in which the application of the algorithm that they describe departs from the original estimation sample in DC[2]. To us, this is an implicit acknowledgment of the importance of our efforts of keeping track of the peculiar features of the format of each edition of the ACN, and of incorporating them in the definition of the estimation sample. Then, the corrigendum presents the estimates obtained from an estimation sample that the Authors describe as obtained when applying the algorithm. We quote here the description of the algorithm provided in the corrigendum:

> "(1) For each qualification slot defined in a head-to-head format, we assign the winner to the treatment group, and the loser to the control group;
>
> (2) For each qualification slot defined in a round-robin group format, we include: In the treatment: for each group, any qualifying team that finished 3 points or less ahead of the team in the group that had the most points while still failing to qualify, as long as the qualifying team in question entered the last matchday of the group with *a nonzero probability of elimination*. In the control: for each group, any non-qualifying team that finished 3 points or less behind the team in the group that had the fewest points while still qualifying, as long as the non-qualifying team in question entered the last matchday of the group with *a nonzero probability of qualification*;
>
> (3) For each qualification slot defined in a "virtual playoff" among top ranked not-directly-qualified teams in a round-robin group, we include: In the treatment: any qualifying team in the "virtual playoff". In the control: any team in the "virtual playoff" that finished 3 points or less behind the team in the "virtual playoff" that had the fewest points while still qualifying." (Depetris-Chauvin et al., 2024, p. 2-3, emphasis added)

The algorithm combines three conditions: (C.1) a team must have an undetermined qualification status before the last match day, (C.2a) a team that qualifies (is eliminated) directly in its own group must be no more than three points above (below) the eliminated (qualified) team with the highest (lowest) number of points in the final table of the same group, and (C.2b) a team that is eliminated in a virtual playoff group must be no more than three points below the team qualifying via the virtual playoff group with the lowest number of points in the table of the virtual playoff group. More precisely, a team qualifying in head-to-head confrontations must satisfy only condition (C.1), a team qualifying or being eliminated in its own group must satisfy (C.1) and (C.2a), while a team qualifying or being eliminated in a virtual playoff group must satisfy only (C.2b), but not (C.1).[20]

[20]Notice that (C.1) is mechanically satisfied before the last match in a head-to-head confrontation, and that the algorithm does not refer to condition (C.2b) for qualified teams in the virtual playoff group, but

Thus, the algorithm introduces sample selection criteria that are dependent on the format of the qualification phase for the ACN,[21] and that also depend on how a team obtained its qualification (or failed to qualify). These differences in the definitions have two key implications: First, the original estimation sample in DC^2 includes, in accordance with the algorithm described in the corrigendum, three teams in the treatment (control) group that won (lost) both matches in head-to-head confrontations in the qualifications for the ACN 2013: Ghana, Mali and Ivory Coast (Malawi, Botswana and Senegal), and thus violate (C.2a), as they ended up 6 points above (below) the other team in the group,[22] as the sample selection criteria depend on the number of teams in a group.[23] Second, (C.1) is not applied in the third part of the algorithm, thus departing from the descriptions of the sample selection criteria included in DC^2, which always refer to the indeterminacy of the qualification status as the criterion determining the inclusion in the estimation sample.[24]

Beyond these initial remarks, there are four key aspects about this algorithm, which are described in detail in Sections M.1-M.4 of the online Appendix. First, the Authors do not discuss what is the relationship between this algorithm and the descriptions of the sample selection criteria included in the original article, and notably whether (C.2a)-(C.2b) are consistent or not with these descriptions. The original article repeatedly states that the estimation sample is built on the basis of an analysis of the situation *before* the last match day, while (C.2a)-(C.2b) assess the situation *after* the last match day, as it uses the points in the final table. Second, the algorithm described in the corrigendum generates neither the estimation sample used in the original article, as the Authors themselves acknowledge at pp. 3-4, nor the estimation sample used in the corrigendum itself. Notably, the corrigendum acknowledges 12 instances in which the original estimation sample in DC^2 is inconsistent

this is immaterial, as (C.2b) is always satisfied for each qualified team.

[21] Notice that the algorithm ignores ignores the real playoff group included in the qualification phase for the ACN 2000 (see Section B.2).

[22] See https://athlet.org/football/can/2013/qualifiers/second-round (last accessed on Dec. 8, 2024).

[23] Notice that a head-to-head format is just a group with two teams.

[24] The original estimation sample in DC^2 includes Cameroon 2012 in the control group, even though (C.1) fails (see Section F.4), as Cameroon was included in a virtual playoff group and it met condition (C.2b); interestingly, the corrigendum states at p. 3 that the inclusion Cameroon 2012 in the control group of the original estimation sample is inconsistent with the algorithm, but this is not the case according to our own reading of the algorithm, which makes no reference, in its third part, to the indeterminacy of the qualification status, which is also not implied by (C.2b).

with the algorithm, but it does not detect additional inconsistencies, even when they occur within the same group where an exclusion error has been identified.[25] Third, the corrigendum refers to a "revised replication package", which is actually not available online, so that we are unable to understand which are the raw data that have been used to assess (C.1), and why the algorithm has been inconsistently applied. Fourth, a consistent application of the algorithm described in the corrigendum confirms 25 of the 41 inclusion and exclusion errors that we have uncovered (see Section 3 above). The remaining cases relate to Cameroon 2012 and to 15 exclusion errors correspond to instances in which (C.2a) is violated: none of these exclusion errors is associated to a probability of qualification or elimination that is lower than those characterizing some of the country-qualification campaign pairs included in the estimation sample in the original article albeit they were virtually qualified or eliminated before the last match day.[26] Thus, (C.2a) is not directly related to the descriptions of the sample selection criteria included in the original article,[27] and it does not effectively restrict the sample to teams with "concrete chances of qualifying" (DC^2, p. 1591).

Our publicly available code, which is based on the descriptions of the sample selection criteria provided in DC^2, implements a definition that is invariant across formats of the qualification phase, and that is independent from the way in which a team obtained its qualification (or failed to qualify), and from the results of the last match day. These natural properties are not satisfied by the algorithm described in the corrigendum, which is currently not publicly available, and which also departs from the original article. Our paper does not discuss whether the choices underlying the algorithm can be justified. It rather points out their inconsistency with the descriptions of the sample selection criteria, and their inability to generate the sample underlying the estimates reported in the original article, upon which

[25]Consider, for instance, the final table of Group 2 (ACN 2004): Guinea 12 (qualified), Niger 9 (eliminated), Ethiopia 9 (eliminated) and Liberia 6 (eliminated); the original estimation sample does not include any of these countries in the estimation sample, even though Guinea and Ethiopia satisfy (C.1), as demonstrated in Sections G.10-G.11, and they also clearly meet (C.2a), as they were just three points apart; the corrigendum acknowledges Guinea 2004 as an exclusion error in the original article at p. 4, but not Ethiopia 2004.

[26]See for instance, Ivory Coast 2008, which satisfies (C.2a) but needed to lose by a margin of at least six goals to be eliminated (see Section A.2.4), or Madagascar 2004, which also satisfies (C.2a), but which needed to win by a margin of at least 17 goals to qualify (see Section A.2.2).

[27]Indeed, the corrigendum itself describes Kenya 2012 and Togo 2015, which were included in the control group in DC^2 even though they violate (C.2a), as "arguably constituting close cases" (p. 3).

its peer-review was based.

6 Concluding remarks

The country-level econometric analysis in DC^2 is exposed to a major measurement error in the application of the sample selection criteria described in the original article and, to a lesser extent, in the definition of the treatment dates. These issues are not fixed by the corrigendum recently published by the Authors (Depetris-Chauvin et al., 2024), which departs from the original article in various respects. The empirical evidence around the effect of a close qualification to the ACN on civil conflict is weaker for various specifications in the main text, once these are estimated with the correct samples and treatment dates.

References

Cederman, L.-E., A. Wimmer, and B. Min (2010): "Why do ethnic groups rebel? New data and analysis," *World Politics*, 62, 87–119.

Clemens, M. A. (2017): "The meaning of failed replications: A review and proposal," *Journal of Economic Surveys*, 31, 326–342.

Depetris-Chauvin, E., R. Durante, and F. Campante (2020): "Building Nations through Shared Experiences: Evidence from African Football," *American Economic Review*, 110, 1572–1602.

——— (2024): "Correction to 'Building Nations through Shared Experiences: Evidence from African Football'," *American Economic Review*, online Corrigendum, 10.1257/aer.20180805, available online since December 3, 2024.

Jürisoo, M. (2024): "International football results from 1872 to 2024," Data: https://github.com/martj42/international_results [Version: 57d91d8, Accessed: April 23, 2024].

Stata (2015): *Stata Data-Management Reference Manual, Release 14*, College Station, TX: StataCorp LLC.

Online Appendix (not for publication)

Online Appendix to Léonie Delpeyrou and Simone Bertoli (2024), "Building Nations through Shared Experiences: Evidence from African Football": a replication of Depetris-Chauvin, Durante and Campante (2020).

Table of Contents

A Descriptions of the estimation sample in DC^2

We report here the four instances in the main text in which DC^2 briefly describe the composition of their estimation sample.

> (Description *A*) "Our approach exploits the quasi-randomness of qualification to the final tournament of the Africa Cup of Nations (ACN) for teams that, *prior to the last round of group stage matches*, could still qualify: for each two teams in the same group that, *going into the very last round of matches*, could still qualify, we attribute the one that actually qualified to the treatment group, and the one that *barely* failed to do so to the control group." (Introduction, p. 1574, emphasis added).

> (Description *B*) "The qualifying stage is usually very competitive, and qualification is often decided only in the last match day based on just a narrow point margin or goal difference. Our strategy consists in [...] identifying teams in the same group that, *until the last match day of the group stage*, were both in a position to qualify, but one of which *barely* did while the other did not." (Section III. Country-Level Analysis, A. Empirical Strategy, p. 1590, emphasis added).

> (Description *C*) "Our key comparison is between countries that *barely* qualified to the ACN finals (our treatment group) and those that did not (our control group). The underlying identification assumption is that if two teams in the same group got to *the last match day* with *concrete chances of qualifying*, which one would actually qualify will be determined by quasi-random circumstances, such as a goal scored in the final minutes of the last match by one side or the other." (Section III. Country-Level Analysis, A. Empirical Strategy, p. 1591, emphasis added).

> (Description *D*) "After *we exclude countries that had already qualified* or *had no shot at qualifying by the time of the last match*, we end up with 49 country-qualification campaign pairs in the treatment group, and 60 in the control group." (Section III. Country-Level Analysis, A. Empirical Strategy, p. 1591, emphasis added).

The four descriptions differ in two key respects: (*i*) the time at which the qualification status should be assessed, and whether (*ii*) an undetermined qualification status is not just a necessary but also a sufficient condition for inclusion in the sample. As far as point (*i*) is concerned, Des. *A*, *B* and *C* clearly state that the qualification status should be assessed before the last match day, while Des. *D* refers to the last match. The two dates differ, for just one team per group, for the 15 groups in the qualification group stages for the ACN 1998 to 2015 with an odd number of teams, as one team does not play on the last match day. With

respect to point (*ii*), Des. *A*, *B* and *C* use an adverb ("barely") or an adjective ("concrete") for the chances of qualification that might suggest that a nonzero chances of qualification or elimination are a necessary but not sufficient condition for inclusion in the estimation sample. Conversely, Des. *D* explicitly states that only the teams that were already certainly qualified or eliminated were excluded from the estimation sample, so that an undetermined qualification status is a necessary and sufficient condition for inclusion in the estimation sample.

As a first step, we analyze the data to clarify the two contradictory elements in the descriptions of the estimation sample included in DC^2.

A.1 Last match or last match day

We describe eight country-qualification campaign pairs corresponding to groups with three teams.[1] These cases, which refer to a team that did not play on the last match day, share the following key feature: the qualification status of the team was fully determined before the last match day, but it was still undetermined before its own last match. Only the one described in Section A.1.1 is included in the estimation sample, while those described in Sections A.1.2-A.1.8 are excluded. Thus, the analysis of the groups with an odd number of teams clearly reveals that the composition of the estimation sample is more (albeit not perfectly) consistent Des. *A*, *B* and *C* (last match day), and much less consistent with Des. *D* (last match) in DC^2.

Included in the original estimation sample in DC^2

A.1.1 Uganda 2004, Group 13, ACN 2004

Uganda was included, with Ghana and Rwanda, in a group with three teams; the first team in the final table obtained a direct qualification for the ACN, while the other two were eliminated; Uganda ended up third in this group, and it was thus eliminated (see https://athlet.org/football/can/2004/qualifiers/group-13). Uganda played its last match on Day 5, and the table before Day 5 was the following: Rwanda 4, Uganda 4, and Ghana 3. The two last two match days were: Ghana vs. Uganda (Day 5), and Rwanda vs. Ghana (Day 6). Uganda drew 1-1 its last match on Day 5, and it was thus eliminated (see Section F.1).

[1] Two more cases correspond to Morocco 2000 and Morocco 2004, both of which are not included by DC^2 in the Excel file `classification.xls` in the replication package, which contains all the countries considered as treated or control, even though they are not used in the estimation sample.

If Uganda had won on Day 5 and, for instance, Rwanda had drawn on Day 6, then final table would have been: Uganda 7, Rwanda 5 and Ghana 4, and thus Uganda would have qualified. So, its qualification status was fully determined before Day 6 but undetermined before Day 5.[2]

Excluded from the original estimation sample in DC[2]

A.1.2 Ghana, Group 1, ACN 1998

Ghana was included, with Angola and Zimbabwe, in a group with three teams; the first and second team in the final table obtained a direct qualification for the ACN, while the third one was eliminated, and Ghana won this group (see https://athlet.org/football/can/1998/qualifiers/group-1). Ghana played its last match on Day 5, and the table before Day 5 was the following: Ghana 4, Zimbabwe 4 and Angola 3. The two last two match days were: Ghana vs. Zimbabwe (Day 5), and Angola vs. Zimbabwe (Day 6). Ghana won 2-1 its last match on Day 5, obtaining a direct qualification. If, for instance, Ghana had lost on Day 5 and Angola had won (as it did) on Day 6, the final table would have been: Zimbabwe 7, Angola 6 and Ghana 4, and thus Ghana would have been eliminated. So, its qualification status was fully determined before Day 6 but undetermined before Day 5. The final table of this group was: Ghana 7 (qualified), Angola 6 (qualified) and Zimbabwe 4 (eliminated).

A.1.3 Cameroon, Group 1, ACN 2000

Cameroon was included, with Eritrea and Mozambique, in a group with three teams; the winner obtained a direct qualification for the ACN, while the runner-up joined a real playoff group with Zimbabwe and the runner-up of Group 5, with the winner of this playoff group also obtaining a qualification. Cameroon ended up first its own group (see https://athlet.org/football/can/2000/qualifiers/group-1). Cameroon played its last match on Day 5, and the table before Day 5 was the following: Cameroon 7, Mozambique 3 and Eritrea 1. The two last two match days were: Cameroon vs. Eritrea (Day 5), and Eritrea vs. Mozambique (Day 6). Cameroon won 1-0 its last match on Day 5, thus obtaining a direct qualification. Cameroon might have been eliminated if it had lost against Eritrea on Day 5, and Eritrea had also won against Mozambique (as it actually did) on Day 6. In this case, Cameroon and Eritrea would have had 7 points each in the final table, and Eritrea would have dominated Cameroon according to the first tiebreaking criterion (number of points in direct confrontations, see Section C.1), as Eritrea and Cameroon had drawn 0-0 on Day 2. In this case, Cameroon would have joined the real playoff group, where it could have been eliminated. Thus, its qualification status was fully determined before Day 6 but undetermined before Day 5.

[2]We can also notice that the Excel file `classification.xls` in the replication package associates to Uganda 2004 the date of the last match day rather than the date of its last match as the treatment date. The actual final table of this group was: Rwanda 7 (qualified), Uganda 5 (eliminated) and Ghana 4 (eliminated).

A.1.4 Burkina Faso, Group 5, ACN 2000

Burkina Faso was included, with Burundi and Senegal, in a group with three teams; the winner obtained a direct qualification for the ACN, while the runner-up joined a real playoff group with Zimbabwe and the runner-up of Group 1, with the winner of this playoff group also obtaining a qualification. Burkina Faso ended up first in its own group (see https://athlet.org/football/can/2000/qualifiers/group-5). Burkina Faso played its last match on Day 5, and the table before Day 5 was the following: Burkina Faso 7, Burundi 3 and Senegal 1. The two last two match days were: Burkina Faso vs. Senegal (Day 5), and Senegal vs. Burundi (Day 6). Burkina Faso drew 1-1 its last match on Day 5, thus obtaining a direct qualification. Burkina Faso might have been eliminated if it had lost against Senegal on Day 5, and Senegal had also won against Burundi (as it actually did) on Day 6. In this case, Burkina Faso and Senegal would have had 7 points each in the final table, and Senegal would have dominated Burkina Faso according to the first tiebreaking criterion (number of points in direct confrontations, see Section C.1), as Senegal and Burkina Faso had drawn 1-1 on Day 2. In this case, Burkina Faso would have joined the real playoff group, where it could have been eliminated. Thus, its qualification status was fully determined before Day 6 but undetermined before Day 5.

A.1.5 Uganda 2002, Group 5, ACN 2002

Uganda was included, with Senegal and Togo, in a group with three teams; the winner and the runner-up obtained a direct qualification for the ACN. Uganda ended up third in this group (see https://athlet.org/football/can/2002/qualifiers/group-5), and so it was eliminated. Uganda played its last match on Day 5, and the table before Day 5 was the following (goal difference between parentheses): Senegal 5 (3), Togo 4 (3) and Uganda 1 (-6). The two last two match days were: Uganda vs. Togo (Day 5), and Togo vs. Senegal (Day 6). Uganda lost 0-3 its last match on Day 5, thus being eliminated. Uganda might have qualified if Togo had lost both matches by a sufficient margin. In this case, Togo and Uganda would have had 4 points each in the final table, and Uganda might have dominated Togo according to the first tiebreaking criterion (total goal difference, see Section C.1). Thus, for Uganda its qualification status was fully determined before Day 6 but undetermined before Day 5.

A.1.6 Nigeria 2004, Group 1, ACN 2004

Nigeria was included, with Angola, Djibouti and Malawi, in a group with four teams; the group was reduced to three teams after Djibouti withdrew, with only the winner obtaining a direct qualification for the ACN, while the other two teams were eliminated. Nigeria won this group (see https://athlet.org/football/can/2004/qualifiers/group-1), and so it was qualified. Nigeria played its last match on Day 5, and the table before Day 5 was the following: Nigeria 7, Malawi 3 and Angola 1. The two last two match days were: Nigeria vs. Angola (Day 5), and Angola vs. Malawi (Day 6). Nigeria drew 2-2 its last match on Day 5, thus being qualified. Nigeria would have been eliminated if Angola had won the last two matches.[3]

[3]Angola won 5-1 on the last match day, and it was leading 0-2 in Nigeria in the second half, before Nigeria managed to draw (see https://www.footballdatabase.eu/en/match/overview/

In this case, Angola and Nigeria would have had 7 points each in the final table, and Angola would have dominated Nigeria according to the first tiebreaking criterion (points in direct confrontations, see Section C.1), as Angola and Nigeria had drawn 0-0 on Day 1. Thus, the qualification status of Nigeria was fully determined before Day 6 but undetermined before Day 5. The actual final table of this group was: Nigeria 8 (qualified), Angola 5 (eliminated) and Malawi 3 (eliminated).

A.1.7 Senegal 2004, Group 8, ACN 2004

Senegal was included, with Gambia, Lesotho and São Tomé-et-Principe, in a group with four teams; the group was reduced to three teams after São Tomé-et-Principe withdrew, with only the winner obtaining a direct qualification for the ACN, while the other two teams were eliminated. Senegal won this group (see https://athlet.org/football/can/2004/qualifiers/group-8), and so it was qualified. Senegal played its last match on Day 5, and the table before Day 5 was the following (goal difference between parentheses): Senegal 7 (3), Gambia 4 (4) and Lesotho 1 (-7). The two last two match days were: Senegal vs. Lesotho (Day 5), and Lesotho vs. Gambia (Day 6). Senegal won 3-0 its last match on Day 5, thus being qualified. Senegal would have been eliminated if it had lost its last match, and Gambia had won on the last latch day. In this case, Gambia and Senegal would have had 7 points each in the final table, and Gambia would have certainly dominated Senegal according to the first tiebreaking criterion (total goal difference, see Section C.1). Thus, the qualification status of Senegal was fully determined before Day 6 but undetermined before Day 5.

A.1.8 Namibia 2004, Group 12, ACN 2004

Namibia was included, with Algeria and Chad, in a group with three teams; the winner obtained a direct qualification for the ACN, while the other two teams were eliminated. Namibia ended up third its own group (see https://athlet.org/football/can/2004/qualifiers/group-12), and so it was eliminated. Namibia played its last match on Day 5, and the table before Day 5 was the following (goal difference between parentheses): Algeria 6 (4), Chad 3 (-2) and Namibia 3 (-2). The two last two match days were: Algeria vs. Namibia (Day 5), and Chad vs. Algeria (Day 6). Namibia lost 1-0 its last match on Day 5, thus being eliminated. Namibia might have qualified by winning against Algeria, and if Chad defeated Algeria on the last match day. In this case, the three teams would have had 6 points each in the final table, and Namibia might have had the best goal difference with adequate results in the last two matches (e.g., defeating Algeria by a margin of three goals, while Chad won by a margin of just one goal). Thus, its qualification status was fully determined before Day 6 but undetermined before Day 5.

1035964-nigeria-angola.)

A.2 Necessary or necessary and sufficient condition

We describe here five qualification-campaign pairs included either in the treated or in the control group of the original estimation sample in DC^2 whose qualification status was virtually determined before the last match day. Thus, these cases are inconsistent with the possible interpretation of Des. A, B and C, where "barely" qualified or eliminated requires having more than just nonzero chances of qualification and elimination, while they are consistent with Des. D in DC^2. Thus, the composition of the original estimation sample clearly reveals that "barely" just refers to a qualification or elimination determined on the last match day, and not before, and "concrete chances" just means nonzero chances.

A.2.1 Ivory Coast, Group 3, ACN 2000

Ivory Coast 2000 belongs to the treatment group in the estimation sample of DC^2. Ivory Coast was included in a group of four teams, with Congo, Mali and Namibia, where the first two teams in the final table obtained a direct qualification, and the other two were eliminated. The table before the last match day was (goals for and goals against between parentheses): Ivory Coast 10 (7 and 2), Mali 8 (5 and 3), Congo 7 (3 and 5) and Namibia 1 (2 and 7). The last match day was: Ivory Coast vs. Mali and Congo vs. Namibia. As the first two tiebreakers in this edition were the total goal difference and the number of goals scored (see C.1), Ivory Coast might have been eliminated if Mali won its away match and if Congo won against Namibia, with the sum of the margins of the two victories being equal to seven goals or more.

A.2.2 Madagascar, Group 10, ACN 2004

Madagascar 2004 belongs to the control group in the estimation sample of DC^2. Madagascar was included in a group of three teams, with Egypt and Mauritius, where the first team in the final table obtained a direct qualification, and the other two were eliminated. The table before the last match day was (goals for and goals against between parentheses): Egypt 9 (14 and 1), Madagascar 6 (2 and 6) and Mauritius 0 (0 and 9). The last match day was: Madagascar vs. Mauritius. As the first two tiebreakers in this edition were the total goal difference and the number of goals scored (see Table C.1), Madagascar could have qualified only by winning against Mauritius by a margin of at least 17 goals. This margin is seven goals larger than the largest margin of victory even observed in a ACN qualifier game, which corresponds to São Tomé-e-Principe vs. Nigeria 0-10, played on June 13, 2022 in the qualification group stage for the ACN 2023 (Jürisoo, 2024).

A.2.3 Gabon, Group 1, ACN 2008

Gabon 2008 belongs to the control group in the estimation sample of DC^2. Gabon was included, with Ivory Coast and Madagascar, in a group of three teams. The first team in the final table obtained a direct qualification, while the other two teams were eliminated. The table before the last match day was (goal

difference between parentheses): Ivory Coast 9 (13), Gabon 6 (1) and Madagascar 0 (-14). The last match day was: Gabon vs. Ivory Coast. The first three tiebreakers in this edition were centered on the results in direct confrontations (see Table C.1). Gabon had lost 5-0 in Ivory Coast on Match Day 2, so it could have won the group only by winning the last match by a margin of at least six goals.[4] This is a margin that exceeds both the largest victory for Gabon and the lagest defeat for Ivory Coast ever observed in ACN qualifiers (Jürisoo, 2024).

A.2.4 Ivory Coast, Group 1, ACN 2008

Ivory Coast 2008 belongs to the treatment group in the estimation sample of DC^2; see Section A.2.3.

A.2.5 Gambia, Group 8, ACN 2008

Gambia 2008 belongs to the control group in the estimation sample of DC^2. Gambia was included in a group of four teams, with Algeria, Cape Verde and Guinea. The first team in the final table obtained a direct qualification, the runner-up joined a virtual playoff group including the runner-ups of Group 2 to 11 (qualifying the first three teams in the virtual playoff), while the other two teams were eliminated. The table before the last match day was (goals for and goals against between parentheses): Guinea 8 (6 and 3), Algeria 8 (5 and 4), Gambia 5 (4 and 5) and Cape Verde 5 (3 and 6). The last match day was: Gambia vs. Algeria and Guinea vs. Cape Verde. The first three tiebreakers in this edition were centered on the results in direct confrontations, while a tie between all four teams would have been broken on the basis of the total goal difference, and the total goals scored (see Table C.1). Gambia had a theoretical possibility to win the group, or to qualify as one of the three best runner-ups via the virtual playoff group. However, both cases were virtually impossible to happen.

Direct qualification: Gambia could have won the group if (i) it won on the last match day by a margin $x_1 \geq 2$, and (ii) Cape Verde won its away match in Guinea by a margin $x_2 \in [\min\{1, 5 - x_1\}, x_1 + 4]$. In this case, all teams would have had 8 points, and Gambia would have had the best total goal difference. Condition (ii) seems, in particular, very unlikely, as, before the last match day of this group, Cape Verde had won a single away match in all the previous editions of the ACN qualifiers (Jürisoo, 2024).[5]

Indirect qualification: Gambia could have joined the virtual playoff group only if (iii) it won on the last match day by a margin $x_1 \geq 2$, and condition (ii) is not satisfied. Gambia would have had 8 points, as the runner-ups of Group 4 and of Group 9 were certain to join the virtual playoff group with more than 8 points (see Table E.5). Thus, Gambia could have, at best, ranked third, the last place in the virtual playoff granting an indirect qualification. The runner-ups of other six groups were certain to join the virtual playoff group with at least 8 points (see Table E.5). In 99.95 percent of the combinations of points of the last match

[4]With a margin of five goals, Ivory Coast would have dominated Gabon either with respect to the third or the fourth tiebreaking criteria, respectively number of away goals in direct confrontations and total goal difference.

[5]Mauritania vs. Cape Verde 0-2 played on September 6, 2002.

day[6], at least another runner-up would have had more than 8 points, and thus Gambia would have been eliminated. In 0.05 percent of the cases, Gambia would have been involved in a tie for the third place with six or seven other teams, qualifying only if it had the best goal difference in this very large set of tied teams. This was, admittedly, almost impossible happen.

[6]This share can be computed by analyzing the cumulative distribution of points of the runner-up in each of the other groups reported in Table E.2 in Section E.

B The qualification group stage for the ACN

B.1 ACN 1998

16 teams participated in the ACN, hosted in Burkina Faso. Burkina Faso (the host country) and South Africa (title holder) were automatically qualified. After a preliminary qualification phase, the remaining 14 places in the final phase were attributed via a qualification group stage. This consisted of seven groups of four teams, each playing against the other twice. Excluding the teams that withdrew or were disqualified[7], 25 teams participated in the qualification group stage. The first and the second team in the final table of each group qualified for the ACN. No playoff for the qualification group stage in this edition.

B.2 ACN 2000

16 teams participated in the ACN, which was initially planned to be hosted in Zimbabwe. Zimbabwe (host country) and South Africa (title holder) were automatically qualified. After a preliminary qualification stage, the remaining 14 places in the final phase were attributed via a qualification group stage. This initially consisted of seven groups of four teams, with the first and the second team in the final table of each group qualified for the ACN. After the groups had been formed (and some matches played), the CAF decided to withdraw the organization of the final phase from Zimbabwe, and it decided to award it jointly to Ghana and Nigeria (source: https://en.wikipedia.org/wiki/2000_African_Cup_of_Nations, last accessed on March 12, 2024). After this change (plus the fact that Sierra Leone withdrew from Group 2) 27 teams participated in the qualification group stage. while Ghana and Nigeria were automatically qualified for the final phase (as host countries), and they were thus withdrawn from Group 1 (Ghana) and Group 5 (Nigeria), with the results of their matches being annulled. After this change, only the first team in the final table of Group 1 and 5 qualified directly for the final phase. The runner-ups of Group 1 and 5 joined a play-off with Zimbabwe, which had not been included in the qualification phase as it was initially awarded an automatic qualification as the host country. The three teams in the playoff group played against each other in two legs, and the first team in the play-off qualified for the final phase.

B.3 ACN 2002

16 teams participated in the ACN, hosted in Mali. Mali (host country) and Cameroon (title holder) were automatically qualified. After a preliminary qualification phase, the remaining 14 places in the final phase were attributed via a qualification group stage. This consisted of seven groups of four teams.[8] Excluding the disqualified team, 27 teams participated in the qualification group stage. The first and the second team

[7] In Group 1, Sudan withdrew after having played one match (which was not counted); in Group 4, the Central African Republic was disqualified after having played one match (which was not counted), and Sierra Leone withdrew after having played two matches.

[8] In Group 5, Guinea was disqualified (results annulled).

in the final table of each group qualified for the ACN. No playoff for the qualification group stage in this edition.

B.4 ACN 2004

16 teams participated in the ACN, hosted in Tunisia. Tunisia (host country) and Cameroon (title holder) were automatically qualified. The remaining 14 places in the final phase were attributed via a qualification group stage. This consisted of 13 groups (Groups 1 to 10 with four teams, and Groups 11 to 13 with three teams). The teams played against each other in two legs. In three groups, all initially included four teams, three teams (Djibouti in Group 1, São Tomé-et-Principe in Group 8, and Guinea-Bissau in Group 10) withdrew from the qualification group stage (results annulled), while Tanzania, which was in Group 3, was disqualified, but the results were not annulled. Excluding the teams that withdrew or were disqualified, 45 teams participated in the qualification group stage. The first team in each group was qualified for the ACN. The final place was assigned to a virtual playoff, which included the runner-ups in the groups with four teams, except Groups 1, 8 and 10, where one team withdrew. The team with the highest number of points in the virtual playoff group qualified for the ACN.

B.5 ACN 2006

16 teams participated in the ACN, hosted in Egypt. Egypt (host country) was automatically qualified, while the title holder (Tunisia) was not automatically qualified to the final phase. Egypt nevertheless took part in the qualifications, as these were also valid for the World Cup 2006 in Germany. After a preliminary qualification phase, the 15 remaining places in the final phase were attributed via a qualification group stage. This consisted of five groups of six teams, each playing against the others in two legs.[9] 30 teams participated in the qualification group stage. As Egypt was admitted to the qualification group stage, three other teams in its group qualified for the final phase of the ACN, i.e., the first three teams excluding Egypt qualified for the ACN in the group which included Egypt. The first three teams in the final table of each group qualified for the ACN, except in Group 3, which included Egypt, where four teams qualified (Egypt plus the other three teams with the highest number of points in the final table). No playoff for the qualification group stage in this edition.

B.6 ACN 2008

16 teams participated in the ACN, hosted in Ghana. The host country (Ghana) was automatically qualified to the final phase. The remaining 15 places were assigned through a qualification group stage. This consisted of 11 groups of four teams, plus a group with three teams (Group 12), each playing against the others in two legs. Djibouti withdrew from Group 1, which was reduced to three teams. Excluding the team that

[9]The same groups also determined the qualification for the World Cup 2006, with the first team in each group qualifying for the World Cup.

withdrew, 46 teams participated in the qualification group stage. The first team in each group qualified for the ACN. The three remaining places were assigned to a virtual playoff, which included the runner-ups in Groups 2 to 11. The three teams with the highest number of points in the playoff group qualified to the ACN. The last match day was September 8 or September 9, 2007 in all groups, except Group 2 (October 13, 2007) and Group 9 (October 12, 2007). This implies the ranking in the (virtual) playoff group was finalized only on October 13, 2007. A match of the second to the last match day in Group 2 (Burundi vs. Egypt) was played on September 9, 2007, when other groups had already played their last match day.

B.7 ACN 2010

16 teams participated in the ACN, hosted in Angola. The host country (Angola) was automatically qualified for the final phase. Angola nevertheless took part in the qualifications, as these were also valid for the World Cup 2010 in South Africa. After a preliminary qualification phase, the 15 places in the final phase were attributed via a qualification group stage, to which Angola failed to be admitted. This consisted of five groups of four teams, each playing against the others in two legs.[10] 20 teams participated in the final group stage, where the first team in each of the five groups obtained qualification for the World Cup 2010. The first three teams in each qualified for the ACN. No playoff for the qualification group stage in this edition.

B.8 ACN 2012

16 teams participated in the ACN, hosted in Equatorial Guinea and Gabon. The host countries (Equatorial Guinea and Gabon) were automatically qualified for the final phase. After a preliminary qualification phase, the remaining 14 places in the final phase were attributed via a qualification group stage. This consisted of 10 groups (Group A to J) of four teams, and a group (Group K) with five teams, each playing against the others in two legs. Mauritania withdrew from Group C before the beginning of the matches. Excluding the team that withdrew, 44 teams participated in the qualification group stage. The first team of each group qualified for the ACN, except for Group K, where the first two teams qualified for the ACN. The two remaining places were assigned to a virtual playoff, which included the runner-ups in Groups A to J. As Group C was reduced to three teams only, the results of the runner-up against the fourth team in each one of the other groups were not considered for the virtual playoff. The two teams with the highest number of (adjusted) points in the playoff group qualified for the ACN.

B.9 ACN 2013

16 teams participated in the ACN, hosted in South Africa. South Africa (the host country) was automatically qualified. After a preliminary qualification phase, the remaining 15 places in the final phase were

[10]The same groups also determined the qualification for the World Cup 2010, with the first team in each group qualifying for the World Cup in South Africa.

attributed through direct confrontations (in two legs). After preliminary rounds, 30 teams participated in the qualification group stage. The winner of each direct confrontation (group) qualified.

B.10 ACN 2015

16 teams participated in the ACN; Morocco should have been the host country, but it demanded (in November 2014) to postpone the final phase to 2016 to an outbreak of Ebola in various African countries; the CAF refused this request, awarding the organization to Equatorial Guinea (on November 14, 2014). Equatorial Guinea (the host country) was automatically qualified (as host country), after having been disqualified from a preliminary phase of the qualification stage, so this change in the host country was, differently from what happened in the ACN 2000, immaterial for the qualification group stage. After a preliminary qualification phase, the remaining 15 places in the final phase were attributed via a qualification group stage. This consisted of seven groups of four teams, each playing against the others in two legs. 28 teams participated in the qualification group stage. The first and the second team in each group qualified for the ACN. The final place was assigned via a virtual playoff, which included the teams ranked third in each of the seven groups. The team with the highest number of points qualified for the ACN.

C Tiebreaking criteria

The set of criteria that are sequentially used in case of a tie between two or more teams in each edition (except 1998) have been retrieved using the Wayback Machine of the Internet Archive (see https://web.archive.org/) applied to the website of the CAF, the African Football Confederation. These criteria are summarized in Table C.1. Our inability to retrieve the information for 1998 is immaterial, as in the only potentially decisive tie (between Algeria and Ivory Coast in Group 2), both the results in direct confrontations and total goal difference would have broken this (potential) tie in favor of Algeria.

Table C.1: Tiebreaking criteria in the qualifications group stage

	1998	2000	2002	2004	2006	2008	2010	2012	2013	2015
All matches										
Goal difference		1	1	1	5	5	5	5		4
Goals scored		2	2	2	6	6	6	6		5
Direct confrontations										
Points		3	3	3	1	1	1	1		1
Goal difference		4	4	4	2	2	2	2	1	2
Goals scored					3	3	3	3	2	
Away goals scored					4	4	4	4	3	3
Additional criteria										
Tiebreaking match		5	5	5						
Random draw					7	7	7	7		6

D Formal definition of the estimation sample

We provide the formal definitions of the estimation sample, treatment and control group, and the treatment date, which apply to all the formats of the qualification group stage of the ten editions of the ACN used in the econometric analysis in DC[2], which are described in the Section B in the online Appendix.

The elements that we use in the definitions are, for each edition of the ACN: (*i*) the format(s) of the groups in the qualification group stage[11], (*ii*) the composition of each group, (*iii*) the points of each team before the last match day of the qualification group stage, (*iv*) the date of each match played in the last match day of all groups, and (*v*) the set of criteria that are sequentially used in case of a tie between two or more teams (see Table C.1 in the online Appendix). We can conveniently analyze all possible combinations of points obtained by each team on the last match day of the qualification group stage, assuming that ties are broken with a random draw. We demonstrate that the resulting definition of the estimation sample is exact if ties are broken first using the goal difference in all matches rather than the results in direct confrontations. For the five editions in which the results in direct confrontations intervene as the first tiebreaker, we verify whether our definition hinges on a tie for which the relative rankings of the teams were already fully determined by the results in the previous match days. This greatly reduces the necessary data collection effort.

Our definition of the estimation sample considers the situation before the first match of the last match day of the group qualification stage that is relevant in the determination of the qualification status of a team. This corresponds to the date of the first match of a group, except when the group also offers access to a virtual playoff group as, in this case, we consider the date of the first match of the last match day of all groups that provide admission to the playoff. The date of the treatment is the date at which the qualification status of a team was eventually determined. This can occur later with respect to the date on which a team played its last match of the group qualification stage.

[11] In the (not so infrequent) case in which the composition of a group is modified, possibly with an ensuing change in the rules that apply to it, our analysis considers the final composition and rules of a group.

D.1 Notation

A set of N_k teams are included in group k, with the qualification group stage of the ACN including a set of G groups. In the ten editions of the ACN that DC[2] focus on, $N_k = 2, ..., 6$.[12] The teams in group k play against each other in two legs; the $N_k(N_k - 1)$ matches distributed over $D(N_k) = N_k(N_k - 1)/O(N_k)$ distinct match days, where $O(N_k)$ is the highest odd number such that $O(N_k) \leq (N_k - 1)$. Each match day $d = 1, .., D(N_k)$ includes $m(N_k) = [O(N_k) + 1]/2$ matches.

We use the index $j = 1, ..., N_k$ for the teams $j \in k$. The ranking $r_j \in R_{N_k} = \{1, ..., N_k\}$ of team j in the final table of group k determines whether team $j \in k$ is directly qualified for the final phase, eliminated or (when a group offers this opportunity) admitted to a playoff group.

$L_k \geq 1$ represents the number of teams in group k that obtain a direct qualification to the final phase of the ACN, i.e., team $j \in k$ is certainly qualified if its rank is not higher than L_k. $\mathbb{I}_k^P$ represents an indicator variable equal to 1 if group k offers access to a playoff group to the team with a rank $r = L_k + 1$, and 0 otherwise[13]; the teams whose rank is $r > (L_k + \mathbb{I}_k^P)$ are certainly eliminated from the competition. When $\mathbb{I}_k^P = 1$, the qualification status of a team $j \in k$ that might have a rank equal to $L_k + 1$ on the basis of the set of possible results of the last match day $D(N_k)$ might be still undetermined, and possibly dependent on the final tables of other groups, or already determined, as explained below. We also define $\mathbb{I}^P \equiv \max_{k \in G} \mathbb{I}_k^P$, i.e., $\mathbb{I}^P = 0$ if and only if the qualification group stage does not include a playoff group.

Without loss of generality, the subscripts are assigned to the various teams in such a way that the matches of the last match day $d = D(N_k)$ are: Team 1 vs. Team 2 if $m(N_k) = 1$; Team 1 vs. Team 2 and Team 3 vs. Team 4 if $m(N_k) = 2$; Team 1 vs. Team 2, Team 3 vs. Team 4 and Team 5 vs. Team 6 if $m(N_k) = 3$.

The set $p \in \{0, 1, 3\}$ gathers the number of points that can be obtained by each team $j \in k$ on each match day d.[14]

[12] N_k is equal to 1 only in the case of a (fictitious) group created in the qualification group stage of the ACN 2000 (see Section B.2 in the online Appendix).

[13] The fictitious group for Zimbabwe in the qualification group stage for the ACN 2000 is the only group for which $L_k = 0$, as this group only offered a (certain) access to a playoff group, i.e., $\mathbb{I}_k^P = 1$.

[14] 3 points are awarded for a victory, 1 point in case of a draw, and 0 for a defeat; a team that does not play on a match day also obtains 0 points.

The vector $\boldsymbol{t}_k$ gives the number of points of each team $j \in k$ before the last match day, i.e., this is the table of points after the match day $d = D(N_k) - 1$. We denote with $\boldsymbol{t}_{lk}$ the points obtained by each team $j \in k$ against a team $l \in k$ in the first $D(N_k) - 1$ match days. Letting N be the highest value of N_k for, with $k = 1, ..., G$, we define a vector $\boldsymbol{t}_k^*$ containing N elements, with $t_{jk}^* = t_{jk}$ if $j \leq N_k$, and 0 otherwise. We can stack all the vectors $\boldsymbol{t}_k^*$ in a $N \times G$ matrix T; this matrix records the points of all teams in all groups before the last match day.

We can keep track of the points obtained by each team on the last match day with the vector $\boldsymbol{p}^k \in p^{m(N_k)}$, where each element $p_j^k \in p$ records the points obtained by the team j in the match day $d = D(N_k)$. As the points obtained by the team playing at home determine the points obtained by the team playing away, we can define a subvector $\boldsymbol{\omega}$ with just $m(N_k)$ elements, which records the points of each team playing at home in the last match day.[15] We use the notation $\boldsymbol{p}^k = p_{N_k}(\boldsymbol{\omega})$ to denote the unique vector $\boldsymbol{p}^k$ that corresponds to any vector $\boldsymbol{\omega} \in \Omega^{m(N_k)}$ for a group with N_k teams.[16]

The set $\Omega^{m(N_k)}$ includes all possible combinations of points $\boldsymbol{\omega}$ in the last match day of group k. The sets $p^{m(N_k)}$ and $\Omega^{m(N_k)}$ include $3^{m(N_k)}$ elements each. We also denote with Ω the set of all possible combinations of points in the last match day all groups $k = 1, ..., G$. We also denote with $\boldsymbol{\omega}_{-k} \in \Omega_{-k}$ a possible combination of points in the last match day in all the groups other than k.

The final table of group k induced by the combination of points $\boldsymbol{\omega} \in \Omega^{m(N_k)}$ in the last match day is the vector:

$$\boldsymbol{t}_k^{D(N_k)}(\boldsymbol{\omega}) = \boldsymbol{t}_k + p_{N_k}(\boldsymbol{\omega})$$

We denote with $\boldsymbol{t}_{lk}^{D(N_k)}(\boldsymbol{\omega}) = \boldsymbol{t}_{lk} + p_{lN_k}(\boldsymbol{\omega})$ the points obtained by each team $j \in k$ against a team $l \in k$, which are equal to 0 for $j = l$.

This final table can include one (or more) ties in terms of points between at least two teams, so that $\boldsymbol{\omega}$ only defines a team-specific correspondence $r_{jk}(\boldsymbol{\omega})$, which gives all possible rankings of team $j \in k$ induced by $\boldsymbol{\omega}$, given $\boldsymbol{t}_k$.

The upper inverse correspondence $r_{jk}^u(r) = \{\boldsymbol{\omega} \in \Omega^{m(N_k)} : r \in r_{jk}(\boldsymbol{\omega})\}$ gives all the combinations of points in the final match day of group k such that team j can have a rank equal to r. $r_{jk}(\boldsymbol{\omega})$ is a singleton if and only if the final table $\boldsymbol{t}_k + p_{N_k}(\boldsymbol{\omega})$ does not include a

[15] For instance, if $m(N_k) = 2$, then $\boldsymbol{\omega} \equiv (p_1\, p_3)'$.

[16] It is necessary to index the function p with the size N_k of group k, as groups of different sizes can have an identical value of $m(N_k)$, and hence vectors $\boldsymbol{\omega}$ of identical size.

tie for team $j \in k$. $r^u_{jk}(r) = \varnothing$ if no combination of results $\boldsymbol{\omega}$ implies that $r \in r_{jk}(\boldsymbol{\omega})$, while $r^u_{jk}(r) = \Omega^{m(N_k)}$ if the ranking of team j in the final table of group j is already determined before the last match day. We define $S^r_k(\boldsymbol{\omega}) \equiv \{j \in k : r \in r_{jk}(\boldsymbol{\omega})\}$ as the set of teams that might have a rank equal to r, given $\boldsymbol{\omega}$ and $\boldsymbol{t}_k$.

The two sets identify the subsets of $\Omega^{m(N_k)}$ such that team j can attain the rank r in the final table of group k with the highest and the lowest number of points. Clearly, $\overline{\Omega}^r_{jk} = \underline{\Omega}^r_{jk} = \varnothing$ if $r^u_{jk}(r) = \varnothing$.

If $\mathbb{I}^P = 1$, we also need to define the indicator variable $\mathbb{I}^V$, which is equal to 1 if the playoff group entails no additional matches, but simply a comparison of the possibly transformed points of the final table of team $j \in S^{L_k+1}_k$, with $\mathbb{I}^P_k = 1$, and 0 otherwise.[17] We define $\mathbb{I}^A_k$ as being equal to 1 if the points of team $j \in S^{L_k+1}_k$, $\mathbb{I}^V_k = 1$, are a transformation of $t^{D(N_k)}_{jk}(\boldsymbol{\omega})$, and 0 otherwise. $\mathbb{I}^P \times (1 - \mathbb{I}^V) = 1$, then the playoff group implies additional matches, in two legs, between the N_P teams admitted to the real playoff group.[18]

The only edition of the group qualification stage for which $\mathbb{I}^P \times \mathbb{I}^V \times \mathbb{I}^A_k = 1$ for some groups $k \in G$ is the one for the ACN 2012. In that edition, we had that $\mathbb{I}^P_k = 1$ only when $N_k = 3, 4$, and $\mathbb{I}^A_k = 1$ only if $N_k = 4$. All the groups k with $\mathbb{I}^P_k = 1$ had $L_k = 1$, i.e., the runner-up was admitted to the (virtual) playoff group. As Group F had $N_k = 3$, the runner-up in all other groups k such that $\mathbb{I}^P_k = 1$ entered the playoff group with an adjusted number of points equal to $t^{2,A}_{jk}(\boldsymbol{\omega}) \equiv t^{D(N_k)}(\boldsymbol{\omega})_{jk} - t^{D(N_k)}_{jlk}(\boldsymbol{\omega})$, where $j \in S^2_k(\boldsymbol{\omega})$, $l \in S^4_k(\boldsymbol{\omega}) \setminus \{k\}$,[19] i.e., only the points obtained against the teams ranked first and third were counted. We define the set $t^{2,A}_{jk}(\boldsymbol{\omega})$ as follows:

$$t^{2,A}_{jk}(\boldsymbol{\omega}) \equiv \left\{ t : t = t^{D(N_k)}_{jk}(\boldsymbol{\omega}) - t^{D(N_k)}_{jlk}(\boldsymbol{\omega}), j \in S^2_k(\boldsymbol{\omega}), l \in S^4_k(\boldsymbol{\omega}) \setminus \{j\} \right\} \tag{D.1}$$

Notice that $t^{2,A}_{jk}(\boldsymbol{\omega})$ is certainly a singleton if $S^4_k(\boldsymbol{\omega})$ is a singleton, while it is not a singleton if $S^4_k(\boldsymbol{\omega})$ includes two or more teams against which team j has obtained a different number of points in direct confrontations.

We define $\overline{t}^{2,A}_{jk}(\boldsymbol{\omega})$ and $\underline{t}^{2,A}_{jk}(\boldsymbol{\omega})$ as the highest and the lowest element in $t^{2,A}_{jk}(\boldsymbol{\omega})$. Similarly, we define $\overline{t}^{2,A}_{jk}$ as the highest value of $\overline{t}^{2,A}_{jk}(\boldsymbol{\omega})$ for $\boldsymbol{\omega} \in r^u_{jk}(L_k + 1)$, and $\underline{t}^{2,A}_{jk}$ as the lowest value

[17] Notice that no edition of the qualification group stage of the ACN included both a real and a virtual playoff, so $\mathbb{I}^V = 0$ corresponds to an edition with a real playoff.

[18] The qualification group stage of the ACN 2000 is the only one to include a real playoff group, with $N_P = 3$.

[19] The condition $l \in S^4_k(\boldsymbol{\omega}) \setminus \{k\}$ allows considering the cases in which $\{2, 4\} \in r_{jk}(\boldsymbol{\omega})$, i.e., $j \in S^4_k(\boldsymbol{\omega})$; in this case, a single random draw determines the identity of the teams ranked second and fourth.

of $\underline{t}^{2,A}_{jk}(\boldsymbol{\omega})$ for $\boldsymbol{\omega} \in r^u_{jk}(L_k+1)$. We also define:

$$\overline{\Omega}^{2,A}_{jk} \equiv \left\{ \boldsymbol{\omega} \in r^u_{jk}(L_k+1) : t^{2,A}_{jk}(\boldsymbol{\omega}) = \overline{t}^{2,A}_{jk} \right\} \tag{D.2}$$

and:

$$\underline{\Omega}^{2,A}_{jk} \equiv \left\{ \boldsymbol{\omega} \in r^u_{jk}(L_k+1) : t^{2,A}_{jk}(\boldsymbol{\omega}) = \underline{t}^{2,A}_{jk} \right\} \tag{D.3}$$

The sets $\overline{\Omega}^{2,A}_{jk}$ and $\underline{\Omega}^{2,A}_{jk}$ gather the combination of points $\boldsymbol{\omega}$ such that team $j \in k$ is admitted to the virtual playoff group respectively with the highest and with the lowest number of adjusted points. We denote with $t^P_{jk}(\boldsymbol{\omega})$ the set of possible points of team $j \in S^{L_k+1}_k(\boldsymbol{\omega})$ in the playoff group if the combination of points of the last match day is $\boldsymbol{\omega} \in \Omega^{m(N_k)}$; this is defined as follows:

$$t^P_{jk}(\boldsymbol{\omega}) \equiv \begin{cases} \{-1\} & \text{if } \mathbb{I}^P_k = 0; \\ \{-1\} & \text{if } \mathbb{I}^P_k = 1,\ \boldsymbol{\omega} \notin r^u_{jk}(L_k+1); \\ \{0\} & \text{if } \mathbb{I}^P_k \times (1-\mathbb{I}^V) = 1,\ \boldsymbol{\omega} \in r^u_{jk}(L_k+1); \\ t^{D(N_k)}_{jk}(\boldsymbol{\omega}) & \text{if } \mathbb{I}^P_k \times \mathbb{I}^V \times (1-\mathbb{I}^A_k) = 1,\ \boldsymbol{\omega} \in r^u_{jk}(L_k+1); \\ t^{2,A}_{jk}(\boldsymbol{\omega}) & \text{if } \mathbb{I}^P_k \times \mathbb{I}^V \times \mathbb{I}^A_k = 1,\ \boldsymbol{\omega} \in r^u_{jk}(L_k+1). \end{cases} \tag{D.4}$$

Notice that Eq. (D.4) fictitiously assigns points in the virtual playoff group to all teams participating in the group qualification stage. In particular, $t^P_{jk}(\boldsymbol{\omega}) = \{-1\}$ for teams included in groups not granting access to the palyoff, or teams for which there is no combination of points on the last match day such their rank is L_k+1. $t^P_{jk}(\boldsymbol{\omega}) = \{0\}$ for teams in groups granting access to a real playoff group, and for which there is at least one combination of points on the last match day such their rank is L_k+1.

We define with $\overline{t}^P_{jk}(\boldsymbol{\omega})$ and $\underline{t}^P_{jk}(\boldsymbol{\omega})$ respectively the highest ad the lowest number of final points of a team $j \in k$ in t^P_{jk}. We also define:

$$t^P_{jk} \equiv \{ t \in t^P_{jk}[r^u_{jk}(L_k+1)] \} \tag{D.5}$$

Then, we define with $\overline{t}^P_{jk}$ and $\underline{t}^P_{jk}$ respectively the highest ad the lowest number in the set t^P_{jk}. Then, we define $\overline{t}^P_k$ as the highest value of $\overline{t}^P_{jk}$ for $j \in S^{L_k+1}_k$, and $\underline{t}^P_k$ as the lowest value of $\underline{t}^P_{jk}$ for $j \in S^{L_k+1}_k$. We also define, for each group $k \in G$, the sets $\overline{\Omega}_k$ such that $j \in S^{L_k+1}_k$ and $\overline{t}^P_{jk} = \overline{t}^P_k$. Similarly, $\underline{\Omega}_k$ is such that the team $j \in S^{L_k+1}_k$ and $\underline{t}^P_{jk} = \underline{t}^P_k$.

We also define:

$$\overline{t}^P(L) \equiv \min_{\text{s.t. } \sum_{k \in G} \mathbb{I}(\overline{t}^P_k \geq n) \leq L} n \in \mathbb{N} \tag{D.6}$$

and:

$$\underline{t}^{P}(L) \equiv \min_{\text{s.t. } \sum_{k \in G} \mathbb{I}(\underline{t}_{k}^{P} \geq n) \geq L} n \in \mathbb{N} \tag{D.7}$$

The variables $\underline{t}^{P}(L)$ and $\overline{t}^{P}(L)$ determine whether a team $j \in k$ such that $j \in S_{k}^{L_k+1}$, i.e., it is in the position to join the playoff group before the last match day in the group qualification phase, can obtain a qualification for the final phase of the ACN via the playoff group (if it joins it with at least $\underline{t}^{P}(L)$ points), or it can be eliminated from the playoff group (if it joins it with less than $\overline{t}^{P}(L)$ points).

We can now define four subsets of $\Omega^{M(N_k)}$, which are specific to each team $j \in k$, and where the two latter sets depend, in general, on T, and can overlap:

$$\Omega^{Q} \equiv \{\boldsymbol{\omega} \in \Omega^{M(N_k)} : \boldsymbol{\omega} \in r_{jk}^{u}(r \leq L_k)\} \tag{D.8}$$

$$\Omega^{E} \equiv \{\boldsymbol{\omega} \in \Omega^{M(N_k)} : \boldsymbol{\omega} \in r_{jk}^{u}(r > L_k + 1)\} \tag{D.9}$$

$$\Omega^{P,Q}(T)(\boldsymbol{\omega}) \equiv \begin{cases} \varnothing & \text{if } \mathbb{I}_{k}^{P} = 0; \\ \boldsymbol{\omega} \in r_{jk}^{u}(L_k + 1) & \text{if } \overline{t}_{jk}^{P}(\boldsymbol{\omega}) \geq \underline{t}^{P}(L) \times \mathbb{I}_{k}^{P} \times \mathbb{I}^{V} = 1 \end{cases} \tag{D.10}$$

$$\Omega^{P,E}(T)(\boldsymbol{\omega}) \equiv \begin{cases} r_{jk}^{u}(L_k + 1) & \text{if } \mathbb{I}_{k}^{P} = 0; \\ \boldsymbol{\omega} \in r_{jk}^{u}(L_k + 1) & \text{if } \underline{t}_{jk}^{P}(\boldsymbol{\omega}) \leq \overline{t}^{P}(L) \times \mathbb{I}_{k}^{P} \times \mathbb{I}^{V} = 1 \end{cases} \tag{D.11}$$

Recall that the definition of the estimation sample is determined before the last match day of the qualification group stage; we do not use the results of the real playoff group to determine the composition of the estimation sample, but just to determine the treatment status of the teams that have been admitted to the playoff group; thus, any $\omega \in r_{jk}^{u}(L_k + 1)$ is included both in $\Omega^{P,Q}(T)$ and in $\Omega^{P,E}(T)$ if $\mathbb{I}_{k}^{P} \times \mathbb{I}^{V} = 0$, as $0 = \underline{t}_{jk}^{P} \leq \overline{t}^{P}(L) \times \mathbb{I}_{k}^{P} \times \mathbb{I}^{V} = 0$ and $0 = \overline{t}_{jk}^{P} \leq \underline{t}^{P}(L) \times \mathbb{I}_{k}^{P} \times \mathbb{I}^{V} = 0$ in this case.

D.2 Definition of the estimation sample and decisive tiebreakers

We can now define the estimation sample using the (admittedly cumbersome but necessary) notation introduced in the previous section, and also identify the cases in which our assumption that random draws used as tiebreakers within groups in the group qualification stage is not immaterial for the inclusion of a team in the estimation sample.

D.2.1 Estimation sample

The indicator variable $\mathbb{I}_{jk}^{U}$ describing whether team $j \in k$ belongs to the estimation sample is equal to:

$$\mathbb{I}_{jk}^{U} = \mathbb{I}_{jk}^{Q} \times \mathbb{I}_{jk}^{E} \tag{D.12}$$

The indicator variables $\mathbb{I}_{jk}^{Q}$ and $\mathbb{I}_{jk}^{E}$ in Eq. (D.12) are defined as:

$$\mathbb{I}_{jk}^{Q} = \begin{cases} 1 & \text{if } \exists! \boldsymbol{\omega} \in \Omega^{Q} \cup \Omega^{P,Q}(T); \\ 0 & \text{otherwise.} \end{cases} \tag{D.13}$$

and:

$$\mathbb{I}_{jk}^{E} = \begin{cases} 1 & \text{if } \exists! \boldsymbol{\omega} \in \Omega^{E} \cup \Omega^{P,E}(T); \\ 0 & \text{otherwise.} \end{cases} \tag{D.14}$$

Letting $\mathbb{I}_{jk}^{FP}$ be equal to 1 if team $j \in k$ actually qualified to the final phase of the ACN, and 0 otherwise, we can rewrite Eq. (D.12) as follows:

$$\mathbb{I}_{jk}^{U} = \begin{cases} \mathbb{I}_{jk}^{E} & \text{if } \mathbb{I}_{jk}^{FP} = 1; \\ \mathbb{I}_{jk}^{Q} & \text{otherwise.} \end{cases} \tag{D.15}$$

D.2.2 Treatment and control group

Team $j \in k$ belongs to the treatment sample if $\mathbb{I}_{jk}^{U} \times \mathbb{I}_{jk}^{FP} = 1$, and to the control group if $\mathbb{I}_{jk}^{U} \times \mathbb{I}_{jk}^{FP} = 0$.

D.2.3 Potentially decisive ties

Consider a team $j \in k$ such that $\mathbb{I}_{jk}^{FP} = 1$; its inclusion in the estimation sample might depend on our assumption that random draws are used as tiebreakers. In particular, this assumption can lead to an incorrect inclusion of a team in the estimation sample. We define $r_{jk}^{*}(\boldsymbol{\omega}) \subseteq r_{jk}(\boldsymbol{\omega})$ the set of possible rankings of team $j \in k$ once we account for the ties within a non-singleton $r_{jk}(\boldsymbol{\omega})$ might have already been determined either before the last match day $D(N_k)$ or given $\boldsymbol{\omega}$. In all cases except the qualification group stage for the ACN 2012 which involved a virtual playoff group with an adjustment of the points of the runner-up for some groups, the following represents a necessary and sufficient condition for our definition of $\mathbb{I}_{jk}^{U}$ in Eq. (D.12) to be correct:

$$\exists! \boldsymbol{\omega} : r_{jk}^{*}(\boldsymbol{\omega}) \cap \Omega^{E} \cup \Omega^{P,E}(T) \neq \varnothing, \mathbb{I}_{jk}^{FP} = 1 \tag{D.16}$$

Similarly, if $\mathbb{I}_{jk}^{FP} = 0$, the following represents a necessary and sufficient condition for our definition of $\mathbb{I}_{jk}^{U}$ in Eq. (D.12) to be correct:

$$\exists! \boldsymbol{\omega} : r_{jk}^{*}(\boldsymbol{\omega}) \cap \Omega^{Q} \cup \Omega^{P,Q}(T) \neq \varnothing, \mathbb{I}_{jk}^{FP} = 0 \tag{D.17}$$

Two sufficient conditions are clearly the following:

$$\exists! \boldsymbol{\omega} : r_{jk}(\boldsymbol{\omega}) \subseteq \Omega^{E} \cup \Omega^{P,E}(T), \mathbb{I}_{jk}^{FP} = 1 \tag{D.18}$$

$$\exists! \boldsymbol{\omega} : r_{jk}(\boldsymbol{\omega}) \subseteq \Omega^{Q} \cup \Omega^{Q,E}(T), \mathbb{I}_{jk}^{FP} = 0 \tag{D.19}$$

The situation is more complicated with the adjustment of the points of the runner-up, as our definition might be incorrect even if $r_{jk}(\boldsymbol{\omega}) = \{L_k + 1\}$, but (i) there is a tie for the fourth position in the group, as this could influence $\bar{t}_{jk}^{P}(\boldsymbol{\omega})$ and $\underline{t}_{jk}^{P}(\boldsymbol{\omega})$,[20] or (ii) there is a tie for second or the fourth position in other groups, as this could influence $\bar{t}^{P}(L)$ and $\underline{t}^{P}(L)$.[21] In this case, we can only state two stronger sufficient conditions, which disregard the virtual playoff group:

$$\exists! \boldsymbol{\omega} : r_{jk}(\boldsymbol{\omega}) \subseteq \Omega^{E}, \mathbb{I}_{jk}^{FP} = 1 \tag{D.20}$$

$$\exists! \boldsymbol{\omega} : r_{jk}(\boldsymbol{\omega}) \subseteq \Omega^{Q}, \mathbb{I}_{jk}^{FP} = 0 \tag{D.21}$$

In all cases in which the sufficient conditions are violated, e.g., a team j that actually qualified might only be eliminated with a combination of points $\boldsymbol{\omega}$ giving rise to a tie for team j, we analyze whether the inclusion of a team in the estimation sample is correct using the actual tiebreaking criteria used in each edition of the qualification group stage (see Section C).

Tiebreakers in the virtual playoff group

Eqs. (D.12) and (D.15) are correct even if ties in the virtual playoff group are not broken with a random draw, if all groups granting access to the virtual playoff group entail no transformation of the points of the team with a rank of $L_k + 1$. Notice that the definition of $\underline{t}^{P}(L)$ ($\bar{t}^{P}(L)$) implies that team $j \in k$ might be able to qualify (be eliminated) only via the playoff group, and with a table of the playoff group which features a decisive tie for team j. In this case, there is at least one $\boldsymbol{\omega} \in \Omega^{P,Q}(T)$ such that $p_{N_k}(\boldsymbol{\omega})$ assigns 3 points (0 points) to team j. As team j can win (lose) its match on the last match day by an arbitrarily wide

[20] Notice that only $\underline{t}_{jk}^{P}(\boldsymbol{\omega})$ matters if $\mathbb{I}_{jk}^{FP} = 1$, and only $\bar{t}_{jk}^{P}(\boldsymbol{\omega})$ matters if $\mathbb{I}_{jk}^{FP} = 0$

[21] Notice that only $\bar{t}^{P}(L)$ matters if $\mathbb{I}_{jk}^{FP} = 1$, and only $\underline{t}^{P}(L)$ matters if $\mathbb{I}_{jk}^{FP} = 0$

margin, it could certainly have a better (worse) goal difference than the other teams with the same number of points in the virtual playoff group.[22] If points are adjusted, then Eq. (D.12) is correct, for any $\boldsymbol{t}_k$, only if we assume that random draws are used as tiebreakers with the virtual playoff group. This follows from the fact that the vector $\boldsymbol{\omega}$ that maximizes the number of adjusted points of team j as a runner-up might not maximize its total number of points as runner-up in the final table of group k. This corresponds to a perverse (but infrequent) situation in which the matrix T_k is such that a team, certainly unable to win its group, can increase its chances of qualification in the virtual playoff group by losing its last match. A matrix T_k corresponding to this case is:

$$T_k = \begin{pmatrix} 0 & 3 & 3 & 2 \\ 0 & 0 & 0 & 2 \\ 3 & 6 & 0 & 3 \\ 2 & 2 & 0 & 0 \end{pmatrix}.$$

With this matrix, we have that $\boldsymbol{t}_k = T_k'1 = (8\ 2\ 12\ 4)'$; in this case, $S_k^1 = \{3\}$, $S_k^2 = \{1\}$, and $S_k^3 = S_k^4 = \{2, 4\}$. If Team 1 wins its last match against Team 2, then $S_k^4 = \{2\}$, and $t_{jk}^{2,A} = [t_{1k} + p_{1,4}(\boldsymbol{\omega})] - [T_{14,k} + p_{1,4}(\boldsymbol{\omega})] = (8+3) - (3+3) = 5$; if Team 1 loses its last match and Team 3 wins, i.e. $\boldsymbol{\omega} = (0\ 3)'$, then $S_k^4 = \{4\}$ and $t_{jk}^{2,A} = t_{1k} - T_{14,k} = 8 - 2 = 6$. In this case, if Team 1 has 6 adjusted points in the virtual playoff group, its goal difference cannot be arbitrarily improved on the last match day, and this might prevent this team from qualifying if $5 < \underline{t}^P(L) = 6$. As no vector $\boldsymbol{t}_k$ that we observe in the data for the qualification group stage of the ACN 2012 has this feature, our definition in Eq. (D.12) is correct even in this case. Similarly, our definition would be incorrect, for the groups with $N_k = 4$ in the qualification group stage of the ACN 2012, if two teams playing against each other in the last match day $D(N_k)$ were already certain to rank respectively $L_k + 1 = 2$ and $N_k = 4$, as the result of this last match would not change the goal difference of their runner-up. As no vector $\boldsymbol{t}_k$ that we observe in the data for the qualification group stage of the ACN 2012 has this feature, this is not an empirically relevant case, and our definition in Eq. (D.12) is correct even in this case.

[22] All groups but one granting access to a virtual playoff group have $N_k = 4$ so that all teams play on the last match day; Group F in the qualification group stage for the ACN 2012 is the only exception here, as $N_k = 3$ and Namibia ($j = 3$), which was not in a position to win the group, might have ended up second without playing on the last match day; however, it would have just had an adjusted number of points equal to 3, which were insufficient to qualify via the virtual playoff group, as $3 < \underline{t}^P(L)$.

E Distribution of points in virtual playoff

We derive the cumulative distribution $F_k(x)$ of the points of the team joining the virtual playoff group for each group k giving access to the playoff group. $F_k(x)$ is defined as the share of combination of points $\boldsymbol{\omega}$ on the last match day such that the team joining the virtual playoff from group k has no more than x points in the table of the playoff group.[23] Tables E.1-E.4 report the cumulative distributions for the four editions of the qualification group stage with a virtual playoff group.[24] This also allows us deriving the values of $\underline{t}^P$ and $\overline{t}^P$ for the each of these editions (see Table E.5).

Table E.1: Cumulative distributions for the ACN 2004

	Points x				
Group	9	10	11	12	13
2	0.667	1	1	1	1
3	0	0.333	1	1	1
4	0.333	0.778	1	1	1
5	0.111	0.667	0.778	1	1
6	0	0.556	0.889	0.889	1
7	0	0.333	0.667	0.667	1
9	0.333	0.778	1	1	1

Table E.5 reveals that, for the ACN 2008, the runner-up of Group C (either Sudan or Tunisia) was certainly qualified for the final phase of the ACN, irrespective of the results on the last match day, as $\underline{t}^P_C = 13 > 12 = \overline{t}^P$ in 2008. Similarly, it also reveals that all the teams potentially admitted to the virtual playoff group for the ACN 2012 from Group E, Group F and Group H had no chance of qualifying for the final phase of the ACN, as $\overline{t}^P_E < \underline{t}^P$ and $\overline{t}^P_F = \overline{t}^P_H = \underline{t}^P$ for 2012. The qualification status of other teams was also fully determined, conditional upon being admitted to the virtual playoff group, before the last match day, but

[23]For the 2012 edition, we thus consider the number of adjusted points for the team joining the virtual playoff group from groups including four teams; if a combination of points $\boldsymbol{\omega}$ results in an undetermined tie for the second and/or the fourth position, we assign an equal weight to all possible values of the adjusted points with which the runner-up could join the virtual playoff group.

[24]Table E.2 accounts for the fact that Burundi and Egypt (Group 2) still had two matches to play after the day (Sept. 8, 2007) in which other groups had played the matches of their last match day.

Table E.2: Cumulative distributions for the ACN 2008

	Points x						
Group	7	8	9	10	11	12	13
2	0.111	0.296	0.556	0.963	1	1	1
3	0	0.333	0.667	0.667	1	1	1
4	0	0	0	0	0	0.333	1
5	0.333	0.667	0.667	1	1	1	1
6	0	0.333	0.667	0.667	1	1	1
7	0	0.333	0.889	0.889	1	1	1
8	0	0.556	0.889	0.889	1	1	1
9	0	0	0.444	0.778	1	1	1
10	0	0.333	0.778	1	1	1	1
11	0	0.222	0.667	0.667	1	1	1

Table E.3: Cumulative distributions for the ACN 2012

	Points x						
Group	3	4	5	6	7	8	9
A	0	0	0.444	0.667	1	1	1
B	0	0.278	0.556	0.611	0.722	0.722	1
C	0	0	0	0.333	0.667	1	1
D	0	0	0.500	0.944	0.944	1	1
E	0.333	0.667	1	1	1	1	1
F	0.333	0.667	0.667	1	1	1	1
G	0	0	0.222	0.778	1	1	1
H	0.167	0.444	0.611	1	1	1	1
I	0	0	0	0	0.667	1	1
J	0	0.056	0.167	0.778	1	1	1

this cannot be read directly from Table E.5.[25]

[25]More precisely, a team that could gain admission to the virtual playoff group with no more than (at least) $\underline{t}^P$ ($\overline{t}^P$) points was certainly eliminated (qualified) from the virtual playoff group.

Table E.4: Cumulative distributions for the ACN 2015

	Points x						
Group	5	6	7	8	9	10	11
1	0	0	0.556	0.889	0.889	1	1
2	0	0.556	0.889	0.889	1	1	1
3	0.333	0.667	0.667	1	1	1	1
4	0	0	0	0.333	0.778	0.889	1
5	0	0	0.444	0.778	1	1	1
6	0.333	0.667	0.667	1	1	1	1
7	0	0.333	0.667	0.667	1	1	1

Table E.5: $\underline{t}^P$ and $\bar{t}^P$ in the editions with virtual playoffs

	ACN 2004		ACN 2008		ACN 2012		ACN 2015	
Group	$\underline{t}_k^P$	$\bar{t}_k^P$	$\underline{t}_k^P$	$\bar{t}_k^P$	$\underline{t}_k^P$	$\bar{t}_k^P$	$\underline{t}_k^P$	$\bar{t}_k^P$
1 (A)					5	7	7	10
2 (B)	9	10	8	11	4	9	6	9
3 (C)	10	11	8	11	6	8	5	8
4 (D)	9	11	12	13	5	8	6	9
5 (E)	9	12	7	10	3	5	7	8
6 (F)	10	13	8	11	3	6	5	8
7 (G)	10	13	8	11	5	7	6	9
8 (H)			8	11	3	6		
9 (I)	9	11	9	11	7	8		
10 (J)			8	10	4	7		
11 (K)			8	11				
L	1		3		2		1	
$\underline{t}^P$	10		8		6		7	
$\bar{t}^P$	14		12		9		10	

F Inclusion errors

We document here each one of the five inclusion and exclusion errors in Table 1.[26] In particular, for each inclusion error, we demonstrate that there was no combination of points on the last match day that would have eliminated (qualified) a team that actually qualified (failed to qualify) for the final phase of the ACN.

F.1 Uganda, Group 13, ACN 2004

$N_{13} = 3$, $L_{13} = 1$, $\mathbb{I}^{V}_{13} = 0$, $\mathbb{I}^{FP}_{UGA,2004} = 0$; table before the last match day (goal difference between parentheses): Rwanda$_1$ 4 (-1), Ghana$_2$ 4 (1), Uganda$_3$ 5 (0); last match day: Rwanda$_1$ vs. Ghana$_2$. With any $\boldsymbol{\omega} \in \Omega$, Uganda would have ranked second, thus being eliminated. In particular, if $\boldsymbol{\omega} = (0)$ ($\boldsymbol{\omega} = (3)$), Ghana (Rwanda) would have won the group with 7 points. If $\omega = (1)$, the three teams would have had all 5 points, with an unchanged goal difference. In this case, Ghana would have won the group, and Uganda would have ranked second, thus being eliminated. The inclusion of Uganda in the control group is, therefore, incorrect.

F.2 South Africa, Group 2, ACN 2006

$N_2 = 6$, $L_2 = 3$, $\mathbb{I}^{P} = 0$, $\mathbb{I}^{FP}_{ZAF,2006} = 1$; table before the last match day: Cape Verde$_1$ 10, Ghana$_2$ 18, South Africa$_3$ 15, DRC$_4$ 15, Uganda$_5$ 7, Burkina Faso$_6$ 12; last match day: Cape Verde$_1$ vs. Ghana$_2$, South Africa$_3$ vs. DRC$_4$, Uganda$_5$ vs. Burkina Faso$_6$. If $\boldsymbol{\omega} = (p\ 0\ 0)'$, then South Africa and Burkina Faso would have had 15 points each. South Africa dominated Burkina Faso with respect to the number of away goals in direct confrontations (the fourth tiebreaking criterion for the ACN 2006, see Table C.1 in the online Appendix), as it had won 2-0 at home, and lost 3-1 away against Burkina Faso. Hence, with $\boldsymbol{\omega} = (p\ 0\ 0)'$, South Africa would have ranked third, and it would have certainly qualified. With all other $\boldsymbol{\omega} \in \Omega^3$, South Africa would have ranked second or third.[27] and hence it would have also qualified. The inclusion of South Africa in the treatment group is, therefore, incorrect.

[26] We do not discuss here one inclusion (Mauritius, Group 4, ACN 2000) appearing in Table A.20 of the online Appendix in DC^2, as Mauritius is not included in the estimation sample, as the ACLED data are not available for them; similarly, we do not describe here the inclusion errors related to North African countries, that appear in the file `classification.xls` in the replication package, but that do not belong to the estimation sample.

[27] South Africa was dominated by Ghana according to the first tiebreaking criterion, as it had obtained 0 points with Ghana, so it would have ranked second with $\boldsymbol{\omega} = (3\ 3\ p)'$.

F.3 Sudan, Group 4, ACN 2008

$N_4 = 4$, $L_4 = 1$, $\mathbb{I}_4^P = 1$, $\mathbb{I}_{SUD,2008}^{FP} = 1$; table before the last match day: Mauritius$_1$ 1, Seychelles$_2$ 3, Sudan$_3$ 12, Tunisia$_4$ 13; last match day: Mauritius$_1$ vs. Seychelles$_2$, Sudan$_3$ vs Tunisia$_4$. With any $\boldsymbol{\omega} \in \Omega^2$, Sudan would have ranked either first or second. In particular, it would have been admitted to the virtual playoff group with either 12 or 13 points. The runner-ups of all other groups could not have been admitted to the virtual playoff group with more than 11 points (see Table E.5). Hence, Sudan would have qualified irrespective of the results of the last match day of all groups. The inclusion of Sudan in the treatment group is, therefore, incorrect.

F.4 Cameroon, Group E, ACN 2012

$N_D = 4$, $L_D = 1$, $\mathbb{I}_D^V = 1$, $\mathbb{I}_{CMR,2012}^{FP} = 0$; table before the last match: DRC$_1$ 7, Cameroon$_2$ 8, Mauritius$_3$ 0, Senegal$_4$ 13; last match day: DRC$_1$ vs. Cameroon$_2$, Mauritius$_3$ vs. Senegal$_4$. Cameroon could not win the group, lagging 5 points behind Senegal. Before the last match day, it only had obtained 2 adjusted points, as it had obtained 6 out of 8 points against Mauritius, which was certain to rank fourth.[28] Even with $\boldsymbol{\omega} = (0\ p)'$, it would have joined the virtual playoff group with just 5 adjusted points, insufficient to qualify (see Table E.5). The inclusion of Cameroon in the control group is, therefore, incorrect.

F.5 Zambia, Group F, ACN 2015

$N_F = 4$, $L_F = 2$, $\mathbb{I}^P = 0$, $\mathbb{I}_{ZMB,2015}^{FP} = 1$; table before the last match day: Niger$_1$ 2, Mozambique$_2$ 5, Zambia$_3$ 8, Cape Verde$_4$ 12; last match day: Niger$_1$ vs. Mozambique$_2$, Zambia$_3$ vs. Cape Verde$_4$. With any $\boldsymbol{\omega} \in \Omega^2$, Zambia would have ranked either first or second, as it had a margin of three points over Mozambique, which was dominated with respect to the first tiebreaking criterion (see Table C.1), as Zambia had obtained 4 points against Mozambique.[29] Hence, Zambia would have qualified irrespective of the results of the last match day. The inclusion of Zambia in the treatment group is, therefore, incorrect.

[28] See https://athlet.org/football/can/2012/qualifiers/group-e (last accessed on April 10, 2024).
[29] See https://fr.athlet.org/football/can/2015/qualifications/groupe-f (last accessed on April 7, 2024).

G Exclusion errors

We document here each one of the 36 exclusion errors in Table 1.[30] For each exclusion error, we provide one or more examples of a combination of points $\boldsymbol{\omega}$ (possibly coupled with specific results in the last match day) contradicting the exclusion of a team from the estimation sample, e.g., for a team that actually qualified, we present a vector $\boldsymbol{\omega}$ that would have eliminated this team.

G.1 Ivory Coast, Group 2, ACN 1998

$N_2 = 4$, $L_2 = 2$, $\mathbb{I}^P = 0$, $\mathbb{I}^{FP}_{CIV,1998} = 1$; table before the last match day (goal difference between parentheses)[31]: Algeria$_1$ 7 (2), Benin$_2$ 2 (-4), Ivory Coast$_3$ 10 (0), Mali$_4$ 9 (2); last match day: Algeria$_1$ vs. Benin$_2$, Ivory Coast$_3$ vs. Mali$_4$. If $\boldsymbol{\omega} = (3\ 0)'$, Mali would have won the group with 12 points, and Algeria and Ivory Coast would have had 10 points each. Algeria would have certainly dominated Ivory Coast with respect to the goal difference with $\boldsymbol{\omega} = (3\ 0)'$, as its final goal difference would have been at least 3, while the final goal difference for the Ivory Coast would have been negative. Algeria also dominated Ivory Coast with respect to the goal difference in direct confrontations (victory 4-1 at home, and loss 2-1 away). Thus, if $\boldsymbol{\omega} = (3\ 0)'$, then Ivory Coast would have been eliminated, as it would have ranked third, irrespective of the set of tiebreaking criteria for this edition. The exclusion of the Ivory Coast from the treatment group is, therefore, incorrect.

G.2 Tanzania, Group 6, ACN 1998

$N_6 = 4$, $L_6 = 2$, $\mathbb{I}^P = 0$, $\mathbb{I}^{FP}_{TZA,1998} = 0$; table before the last match day: Liberia$_1$ 5, Tanzania$_2$ 5, DRC$_3$ 6, Togo$_4$ 10; last match day: Liberia$_1$ vs. Tanzania$_2$, DRC$_3$ vs. Togo$_4$. If, for instance, $\boldsymbol{\omega} = (0\ 0)'$ Tanzania would have ranked second. Hence, Tanzania was in a position to qualify before the last match day. The exclusion of Tanzania from the control group is, therefore, incorrect. The final table of this group is: Togo 10 (qualified), DRC 9 (qualified), Liberia 8 (eliminated) and Tanzania 5 (eliminated).

G.3 Eritrea, Group 1, ACN 2000

$N_1 = 3$, $L_1 = 1$, $\mathbb{I}^P_1 = 1$, $\mathbb{I}^{FP}_{ERI,2000} = 0$; Eritrea was the runner-up in the final table of this group, thus obtaining admission to a real playoff group. Hence, Eritrea could qualify before the last match day. The exclusion of Eritrea from the control group is, therefore, incorrect. The final table of this group was: Cameroon 10 (qualified), Eritrea 4 (admitted to the real playoff group), and Mozambique 3 (eliminated).

[30] We do not discuss two exclusion errors (Cape Verde, Group 5, ACN 2004 and Group 8, ACN 2008) that appear in Table A.20 of the online Appendix in DC^2, as Cape Verde is not included in the estimation sample (because of missing ACLED data); similarly, we do not describe here the exclusion errors related to North African countries, that appear in the file `classification.xls` in the replication package, but that do not belong to the estimation sample.

[31] See https://athlet.org/football/can/1998/qualifiers/group-2 (last accessed on April 10, 2024).

G.4 Mozambique, Group 1, ACN 2000

$N_1 = 3$, $L_1 = 1$, $\mathbb{I}_1^P = 1$, $\mathbb{I}_{MOZ,2000}^{FP} = 0$; table before the last match day: Eritrea$_1$ 1, Mozambique$_2$ 3, Cameroon$_3$ 10, last match day: Eritrea$_1$ vs. Mozambique$_2$. If $\boldsymbol{\omega} = (p)$, with $p \neq 3$, Mozambique would have ranked second, thus gaining to the real playoff group. The exclusion of Mozambique from the control group is, therefore, incorrect. The final table of this group was: Cameroon 10 (qualified), Eritrea 4 (admitted to the real playoff group), and Mozambique 3 (eliminated).

G.5 Burundi, Group 5, ACN 2000

$N_1 = 3$, $L_1 = 1$, $\mathbb{I}_1^P = 1$, $\mathbb{I}_{BDI,2000}^{FP} = 0$; table before the last match day: Senegal$_2$ 2, Burundi$_3$ 3, Burkina Faso$_3$ 8, last match day: Senegal$_1$ vs. Burundi$_2$. If $\boldsymbol{\omega} = (p)$, with $p \neq 3$, Burundi would have ranked second, thus gaining to the real playoff group. The exclusion of Burundi from the control group is, therefore, incorrect. The final table of this group was: Burkina Faso 8 (qualified), Senegal 5 (admitted to the real playoff group), and Burundi 3 (eliminated).

G.6 Senegal, Group 5, ACN 2000

$N_5 = 3$, $L_5 = 1$, $\mathbb{I}_5^P = 1$, $\mathbb{I}_{SEN,2000}^{FP} = 1$; Senegal was the runner-up in the final table of this group, thus obtaining admission to a real playoff group. Hence, Senegal could have been eliminated before the last match day. The exclusion of Senegal from the treatment group is, therefore, incorrect. The final table of this group was: Burkina Faso 8 (qualified), Senegal 5 (admitted to the real playoff group), and Burundi 3 (eliminated).

G.7 Zimbabwe, fictitious group, ACN 2000

$\mathbb{I}_{ZWE,2000}^{FP} = 0$; when the CAF withdrew the organization of the ACN 2000, it granted Zimbabwe access to a playoff group with the runner-ups of Groups 1 and 5. Hence, Zimbabwe could qualify before the last match day of the group qualification stage. The exclusion of Zimbabwe from the control group is, therefore, incorrect.

G.8 Namibia, Group 1, ACN 2002

$N_1 = 4$, $L_1 = 2$, $\mathbb{I}^P = 0$, $\mathbb{I}_{NMB,2002}^{FP} = 0$; table before the last match day: Namibia$_1$ 5, Nigeria$_2$ 11, Madagascar$_3$ 5, Zambia$_4$ 5; last match day: Namibia$_1$ vs. Nigeria$_2$, Madagascar$_3$ vs. Zambia$_4$. If, for instance, $\boldsymbol{\omega} = (3\ 1)'$, Namibia would have ranked second, obtaining a qualification for the final phase of the CAN. The exclusion of Namibia from the control group is, therefore, incorrect.

G.9 Kenya, Group 3, ACN 2002

$N_3 = 4$, $L_3 = 2$, $\mathbb{I}^P = 0$, $\mathbb{I}_{KEN,2002}^{FP} = 0$; table before the last match day: Morocco$_1$ 10, Gabon$_2$ 5, Tunisia$_3$ 5, Kenya$_4$ 6; last match day: Morocco$_1$ vs. Gabon$_2$, Tunisia$_3$ vs. Kenya$_4$. If, for instance, $\boldsymbol{\omega} = (p\ 0)'$, Kenya

would have ranked second, obtaining a qualification for the final phase of the CAN. The exclusion of Kenya from the control group is, therefore, incorrect.

G.10 Ethiopia, Group 2, ACN 2004

$N_2 = 4$, $L_2 = 1$, $\mathbb{I}_2^V = 1$, $\mathbb{I}_{ETH,2004}^{FP} = 0$; table before the last match day: Niger_1 6, Liberia_2 6, Guinea_3 9, Ethiopia_4 9; last match day: Niger_1 vs. Liberia_2, Guinea_3 vs. Ethiopia_4. If, for instance, $\boldsymbol{\omega} = (p\ 0)'$, Ethiopia would have ranked first, obtaining a qualification for the final phase of the CAN. The exclusion of Ethiopia from the control group is, therefore, incorrect. The final table of this group was: Guinea 12 (qualified), Niger 9 (eliminated), Ethiopia 9 (eliminated) and Liberia 6 (eliminated).

G.11 Guinea, Group 2, ACN 2004

$N_2 = 4$, $L_2 = 1$, $\mathbb{I}_2^V = 1$, $\mathbb{I}_{GIN,2004}^{FP} = 1$; table before the last match day: Niger_1 6, Liberia_2 6, Guinea_3 9, Ethiopia_4 9; last match day: Niger_1 vs. Liberia_2, Guinea_3 vs. Ethiopia_4. If, for instance, $\boldsymbol{\omega} = (3\ 0)'$, Guinea would have ranked third with adequate results in the two matches to ensure that Niger had a better goal difference than Guinea, with the ensuing elimination of Guinea from the final phase of the CAN. The exclusion of Guinea from the treatment group is, therefore, incorrect. The final table of this group was: Guinea 12 (qualified), Niger 9 (eliminated), Ethiopia 9 (eliminated) and Liberia 6 (eliminated).

G.12 Sudan, Group 3, ACN 2004

$N_3 = 4$, $L_3 = 1$, $\mathbb{I}_3^V = 1$, $\mathbb{I}_{SUD,2004}^{FP} = 0$; table before the last match day (goal difference between parentheses): Benin_1 10 (5), Zambia_2 11, Sudan_3 7 (0), Tanzania_4 0; last match day: Benin_1 vs. Zambia_2, Sudan_3 vs. Tanzania_4. If $\boldsymbol{\omega} = (0\ 3)'$, Benin and Sudan would have had both 10 points, and Sudan would have ranked second, thus joining the virtual playoff group with a chance of obtaining an indirect qualification (see Table E.5), if the sum of the margins of the victory for Sudan and of the defeat for Zambia on the last match day, was equal to at least six goals. This follows from the fact that the first tiebreaking criterion was the total goal difference for this edition of the ACN (see Table C.1). The exclusion of Sudan from the control group is, therefore, incorrect. The final table of this group was: Benin 13 (qualified), Zambia 11 (eliminated), Sudan 10 (eliminated), Tanzania 0 (eliminated).

G.13 Burkina Faso, Group 4, ACN 2004

$N_4 = 4$, $L_4 = 1$, $\mathbb{I}_4^V = 1$, $\mathbb{I}_{BFA,2004}^{FP} = 1$; table before the last match day: CAR_1 2, Burkina Faso_2 10, Congo_3 8, Mozambique_4 7; last match day: CAR_1 vs. Burkina Faso_2, Congo_3 vs. Mozambique_4. If, for instance, $\boldsymbol{\omega} = (3\ 3)'$, Burkina Faso would have been the runner-up, thus joining the virtual playoff group with 10 points, which were not sufficient to obtain a certain qualification via the virtual playoff group (see Table E.5). The exclusion of Burkina Faso from the treatment group is, therefore, incorrect.

G.14 Congo, Group 4, ACN 2004

$N_4 = 4$, $L_4 = 1$, $\mathbb{I}_4^V = 1$, $\mathbb{I}_{COD,2004}^{FP} = 0$; table before the last match day: CAR_1 2, Burkina Faso_2 10, Congo_3 8, Mozambique_4 7; last match day: CAR_1 vs. Burkina Faso_2, Congo_3 vs. Mozambique_4. If, for instance, $\boldsymbol{\omega} = (3\ 3)'$, Congo would have ranked first, thus obtaining a direct qualification for the final phase of the CAN. The exclusion of Congo from the control group is, therefore, incorrect.

G.15 Mozambique, Group 4, ACN 2004

$N_4 = 4$, $L_4 = 1$, $\mathbb{I}_4^V = 1$, $\mathbb{I}_{MOZ,2004}^{FP} = 0$; table before the last match day (goal difference between parentheses): CAR_1 2 (-3), Burkina Faso_2 10 (7), Congo_3 8 (1), Mozambique_4 7 (-5); last match day: CAR_1 vs. Burkina Faso_2, Congo_3 vs. Mozambique_4. If, for instance, $\boldsymbol{\omega} = (3\ 0)'$, Burkina Faso and Mozambique would have both had 10 points; Mozambique would have either ranked first[32], thus obtaining a direct qualification or second, thus gaining admission to the virtual playoff group; 10 points were potentially sufficient to qualify via the virtual playoff group (see Table C.1). The exclusion of Mozambique from the control group is, therefore, incorrect.

G.16 Eswatini, Group 9, ACN 2004

$N_9 = 4$, $L_9 = 1$, $\mathbb{I}_9^V = 1$, $\mathbb{I}_{SWZ,2004}^{FP} = 0$; table before the last match day: Libya_1 7, Eswatini_2 8, Botswana_3 2, DRC_4 10; last match day: Libya_1 vs. Eswatini_2, Botswana_3 vs. DRC_4. If, for instance, $\boldsymbol{\omega} = (0\ 3)'$, Eswatini would have ranked first, thus obtaining a direct qualification for the final phase of the CAN. The exclusion of Eswatini from the control group is, therefore, incorrect. The final table of this group was: DRC 11 (qualified), Libya 10 (eliminated), Eswatini 8 (eliminated) and Botswana 3 (eliminated).

G.17 Ghana, Group 13, ACN 2004

$N_1 3 = 3$, $L_1 3 = 1$, $\mathbb{I}_1^V 3 = 0$, $\mathbb{I}_{GHA,2004}^{FP} = 0$; table before the last match day: Rwanda_1 4, Ghana_2 4, Uganda_3 5; last match day: Rwanda_1 vs. Ghana_2. If, for instance, $\omega = (0)$, Ghana would have won the group, thus obtaining a direct qualification for the final phase of the CAN. The exclusion of Ghana from the control group is, therefore, incorrect.

G.18 Botswana, Group 2, ACN 2008

$N_2 = 4$, $L_2 = 1$, $\mathbb{I}_2^V = 1$, $\mathbb{I}_{BWA,2008}^{FP} = 0$; table before the last match day: Egypt_1 9, Botswana_2 7, Mauritania_3 4, Burundi_4 7; last match day: Egypt_1 vs. Botswana_2, Mauritania_3 vs. Burundi_4. If, for instance, $\boldsymbol{\omega} = (0\ 1)'$, Botswana would have won the group, thus obtaining a direct qualification for the final phase of the CAN. The exclusion of Botswana from the control group is, therefore, incorrect.

[32]This occurs if the sum of the margin of the victory of CAR and Mozambique exceeded 12 goals, as the first tiebreaking criterion in 2004 was the total goal difference (see Table E.5).

G.19 Burundi, Group 2, ACN 2008

$N_2 = 4$, $L_2 = 1$, $\mathbb{I}_2^V = 1$, $\mathbb{I}_{BUR,2008}^{FP} = 0$; table before the last match day:[33] Egypt$_1$ 9, Botswana$_2$ 7, Mauritania$_3$ 4, Burundi$_4$ 7; last match day: Egypt$_1$ vs. Botswana$_2$, Mauritania$_3$ vs. Burundi$_4$. If, for instance, $\boldsymbol{\omega} = (1\ 0)'$, Burundi would have either won the group, thus obtaining a direct qualification for the final phase of the CAN, or be the runner-up with 10 points, depending on the margin of its victory in the last match day. 8 points were potentially sufficient to qualify via the virtual playoff group (see Table E.5). The exclusion of Burundi from the control group is, therefore, incorrect.

G.20 Tanzania, Group 7, ACN 2008

$N_7 = 4$, $L_7 = 1$, $\mathbb{I}_7^V = 1$, $\mathbb{I}_{TZA,2008}^{FP} = 0$; table before the last match day: Senegal$_1$ 8, Burkina Faso$_2$ 4, Tanzania$_3$ 8, Mozambique$_4$ 6; last match day: Senegal$_1$ vs. Burkina Faso$_2$, Tanzania$_3$ vs. Mozambique$_4$. If, for instance, $\boldsymbol{\omega} = (p\ 3)'$, Tanzania would have either won the group, thus obtaining a direct qualification for the final phase of the CAN, or be the runner-up with 11 points, depending on the result of the other match, and the margins of victory in two matches (if $p = 3$). 11 points were potentially sufficient to qualify via the virtual playoff group (see Table E.5). The exclusion of Tanzania from the control group is, therefore, incorrect. The final table of this group was: Senegal 11 (qualified), Mozambique 9 (eliminated), Tanzania 8 (eliminated) and Burkina Faso 4 (eliminated).

G.21 Mali, Group 9, ACN 2008

$N_9 = 4$, $L_9 = 1$, $\mathbb{I}_9^V = 1$, $\mathbb{I}_{MLI,2008}^{FP} = 1$; table before the last match day: Togo$_1$ 9, Mali$_2$ 9, Sierra Leone$_3$ 1, Benin$_4$ 8; last match day: Togo$_1$ vs. Mali$_2$, Sierra Leone$_3$ vs. Benin$_4$. If $\boldsymbol{\omega} = (3\ 0)'$, Mali would have ranked third, thus being eliminated. The exclusion of Mali from the treatment group is, therefore, incorrect. The final table of this group was: Mali 12 (qualified), Benin 11 (qualified via the virtually playoff group), Togo 9 (eliminated) and Sierra Leone 1 (eliminated).

G.22 Togo, Group 9, ACN 2008

$N_9 = 4$, $L_9 = 1$, $\mathbb{I}_9^V = 1$, $\mathbb{I}_{TGO,2008}^{FP} = 0$; table before the last match day: Togo$_1$ 9, Mali$_2$ 9, Sierra Leone$_3$ 1, Benin$_4$ 8; last match day: Togo$_1$ vs. Mali$_2$, Sierra Leone$_3$ vs. Benin$_4$. If $\boldsymbol{\omega} = (3\ p)'$, Togo would have won the group, thus gaining a direct qualification. The exclusion of Togo from the control group is, therefore,

[33] The match between Burundi and Egypt of the fifth match day was played on Sept. 9, 2007 (see https://athlet.org/football/can/2008/qualifiers/group-2, last accessed on May 4, 2024), thus after other groups had played the matches of their last match day; the table that we report here includes the result of this match (0-0); this is immaterial, given that Burundi could have obtained a direct qualification even after having drawn against Egypt (and, hence, *a fortiori*, if we consider that it still had two matches to play at the date at which its qualification status is assessed).

incorrect. The final table of this group was: Mali 12 (qualified), Benin 11 (qualified via the virtual playoff group), Togo 9 (eliminated) and Sierra Leone 1 (eliminated).

G.23 Ethiopia, Group 10, ACN 2008

$N_{10} = 4$, $L_{10} = 1$, $\mathbb{I}^{V}_{10} = 1$, $\mathbb{I}^{FP}_{ETH,2008} = 0$; table before the last match day: DRC$_1$ 8, Libya$_2$ 7, Ethiopia$_3$ 6, Namibia$_4$ 7; last match day: DRC$_1$ vs. Libya$_2$, Ethiopia$_3$ vs. Namibia$_4$. If $\boldsymbol{\omega} = (p\ 3)'$, Ethiopia would have ranked second[34], thus obtaining access to the virtual playoff group with 9 points. 9 points were potentially sufficient to qualify via the virtual playoff group (see Table E.5). The exclusion of Ethiopia from the control group is, therefore, incorrect.

G.24 Congo, Group 11, ACN 2008

$N_{11} = 4$, $L_{11} = 1$, $\mathbb{I}^{V}_{11} = 1$, $\mathbb{I}^{FP}_{COD,2008} = 0$; table before the last match day: South Africa$_1$ 11, Zambia$_2$ 8, Chad$_3$ 1, Congo$_4$ 6; last match day: South Africa$_1$ vs. Zambia$_2$, Chad$_3$ vs. Congo$_4$. If $\boldsymbol{\omega} = (3\ 0)'$, Congo would have ranked second, thus obtaining access to the virtual playoff group with 9 points. 9 points were potentially sufficient to qualify via the virtual playoff group (see Table E.5). The exclusion of Congo from the control group is, therefore, incorrect.

G.25 Zambia, Group 11, ACN 2008

$N_{11} = 4$, $L_{11} = 1$, $\mathbb{I}^{V}_{11} = 1$, $\mathbb{I}^{FP}_{ZMB,2008} = 1$; table before the last match day: South Africa$_1$ 11, Zambia$_2$ 8, Chad$_3$ 1, Congo$_4$ 6; last match day: South Africa$_1$ vs. Zambia$_2$, Chad$_3$ vs. Congo$_4$. If $\boldsymbol{\omega} = (3\ 0)'$, Zambia would have ranked third, thus being directly eliminated. The exclusion of Zambia from the treatment group is, therefore, incorrect. The final table of this group was: Zambia 11 (qualified), South Africa 11 (qualified via the virtual playoff group), Congo 7 (eliminated) and Chad 1 (eliminated).

G.26 Togo, Group A, ACN 2010

$N_A = 4$, $L_A = 3$, $\mathbb{I}^{P} = 0$, $\mathbb{I}^{FP}_{TGO,2010} = 1$; table before the last match day: Togo$_1$ 5, Gabon$_2$ 9, Morocco$_3$ 3, Cameroon$_4$ 10; last match day: Togo$_1$ vs. Gabon$_2$, Morocco$_3$ vs. Cameroon$_4$. If $\boldsymbol{\omega} = (0\ 3)'$, Togo would have ranked fourth, thus being eliminated. The exclusion of Togo from the treatment group is, therefore, incorrect.

G.27 Kenya, Group B, ACN 2010

$N_B = 4$, $L_B = 3$, $\mathbb{I}^{P} = 0$, $\mathbb{I}^{FP}_{KEN,2010} = 0$; table before the last match day: Mozambique$_1$ 4, Tunisia$_2$ 11, Kenya$_3$ 3, Nigeria$_4$ 9; last match day: Mozambique$_1$ vs. Tunisia$_2$, Kenya$_3$ vs. Nigeria$_4$. If $\boldsymbol{\omega} = (0\ 3)'$, Kenya

[34] If $\boldsymbol{\omega} = (1\ 3)'$, both DRC and Ethiopia would have had 9 points, but DRC dominated Ethiopia in direct confrontations.

would have ranked third, thus obtaining a direct qualification. The exclusion of Kenya from the control group is, therefore, incorrect.

G.28 Mozambique, Group B, ACN 2010

$N_B = 4$, $L_B = 3$, $\mathbb{I}^P = 0$, $\mathbb{I}^{FP}_{MOZ,2010} = 1$; table before the last match day: Mozambique$_1$ 4, Tunisia$_2$ 11, Kenya$_3$ 3, Nigeria$_4$ 9; last match day: Mozambique$_1$ vs. Tunisia$_2$, Kenya$_3$ vs. Nigeria$_4$. If $\boldsymbol{\omega} = (0\ 3)'$, Mozambique would have ranked fourth, thus being eliminated. The exclusion of Mozambique from the treatment group is, therefore, incorrect.

G.29 Zambia, Group C, ACN 2012

$N_C = 4$, $L_C = 1$, $\mathbb{I}^V_C = 1$, $\mathbb{I}^{FP}_{ZMB,2012} = 1$; table before the last match day: Zambia$_1$ 12, Libya$_2$ 11, Mozambique$_3$ 4, Comores$_4$ 1; last match day: Zambia$_1$ vs. Libya$_2$, Mozambique$_3$ vs. Comores$_4$. If $\boldsymbol{\omega} = (0\ p)'$, with $p \neq 0$, Zambia would have ranked second, gaining access to the virtual playoff group with 6 adjusted points (as it had obtained 6 points against Comores, ranked fourth in the group).[35] 6 points were insufficient to guarantee qualification in the virtual playoff group (see Table E.5), so Zambia might have been eliminated. The exclusion of Zambia from the treatment group is, therefore, incorrect. The final table of this group was: Zambia 13 (qualified), Libya 12 (qualified via the virtual playoff group), Mozambique 7 (eliminated) and Comoros 1 (eliminated).

G.30 Tanzania, Group D, ACN 2012

$N_D = 4$, $L_D = 1$, $\mathbb{I}^V_D = 1$, $\mathbb{I}^{FP}_{TZA,2012} = 0$; table before the last match day (goal difference between parentheses): Morocco$_1$ 8 (4), Tanzania$_2$ 5 (-1), Algeria$_3$ 5 (-5), CAR$_4$ 8 (2); last match day: Morocco$_1$ vs. Tanzania$_2$, Algeria$_3$ vs. CAR$_4$.[36] If $\boldsymbol{\omega} = (0\ 3)'$, all teams would have had 8 points in the final table. The first decisive tiebreaker is the goal difference in direct confrontations (see Table C.1), which coincides, in this case, with the total goal difference. If, for instance, Tanzania had won with a margin of (i) at least three goals, and (ii) larger than the margin of the victory of Algeria, then the final table would have been: Morocco$_1$ 8 (≤ -1), Tanzania$_2$ 8 (≥ 2), Algeria$_3$ 8 (< 2), CAR$_4$ 8 (≤ 1), and Tanzania would have won the group, obtaining a direct qualification. The exclusion of Tanzania from the treatment group is, therefore, incorrect. The final table of this group was: Morocco 11 (qualified), CAR 8 (eliminated), Algeria 8 (eliminated), Tanzania 5 (eliminated).

G.31 Burundi, Group H, ACN 2012

$N_H = 4$, $L_H = 1$, $\mathbb{I}^V_H = 1$, $\mathbb{I}^{FP}_{BDI,2012} = 0$; table before the last match day: Ivory Coast$_1$ 15, Burundi$_2$ 5, Benin$_3$ 5, Rwanda$_4$ 3; last match day: Ivory Coast$_1$ vs. Burundi$_2$, Benin$_3$ vs. Rwanda$_4$. If $\boldsymbol{\omega} = (0\ 0)'$, then

[35] See https://athlet.org/football/can/2012/qualifiers/group-c (last accessed on April 7, 2024).
[36] See https://athlet.org/football/can/2012/qualifiers/group-d (last accessed on April 7, 2024).

Burundi would have ranked second with 8 points, thus gaining access to the virtual playoff group, and Benin fourth. As Burundi had obtained 2 points against Benin[37], it would have had $8 - 2 = 6$ adjusted points in the virtual playoff group. 6 points were potentially sufficient to qualify in the virtual playoff group (see Table E.5). The exclusion of Burundi from the control group is, therefore, incorrect.[38]

G.32 Ghana, Group I, ACN 2012

$N_I = 4$, $L_I = 1$, $\mathbb{I}_I^V = 1$, $\mathbb{I}_{GHA,2012}^{FP} = 1$; table before the last match day: Eswatini$_1$ 0, Congo$_2$ 3, Sudan$_3$ 13, Ghana$_4$ 13; last match day: Eswatini$_1$ vs Congo$_2$, Sudan$_3$ vs. Ghana$_4$. If $\boldsymbol{\omega} = (p\ 3)'$, then Ghana would have ranked second with 13 points, thus gaining access to the virtual playoff group, and either Eswatini or Congo would have ranked fourth. As Ghana had obtained 6 points against both Eswatini or Congo[39], it would have $13 - 6 = 7$ adjusted points in the virtual playoff group. 7 points were potentially insufficient to qualify in the virtual playoff group (see Table E.5). The exclusion of Ghana from the treatment group is, therefore, incorrect.The final table of this group was: Ghana 16 (qualified), Sudan 13 (qualified via the virtual playoff group), Congo 6 (eliminated) and Eswatini 0 (eliminated).

G.33 Ethiopia, Group B, ACN 2015

$N_B = 4$, $L_B = 2$, $\mathbb{I}_B^V = 1$, $\mathbb{I}_{ETH,2015}^{FP} = 0$; table before the last match day: Mali$_1$ 6, Algeria$_2$ 15, Ethiopia$_3$ 3, Malawi$_4$ 6; last match day: Mali$_1$ vs. Algeria$_2$, Ethiopia$_3$ vs. Malawi$_4$. If $\boldsymbol{\omega} = (0\ 3)'$, Algeria would have won the group with 18 points, with Ethiopia, Mali and Malawi having 6 points each. With $\boldsymbol{\omega} = (0\ 3)'$, these three teams would have obtained 6 points each in direct confrontations[40], and the first potentially decisive tiebreaker is the goal difference in direct confrontations (see Table C.1). Before the last match day, the three teams had the following goal difference (excluding the results with Algeria): Ethiopia -2, Malawi 1, Mali 1. If Ethiopia defeated Malawi in the last match day with a margin of at least four goals, the three teams would have had the following final goal difference in direct confrontations: Ethiopia ≥ 2, Malawi ≤ -3, Mali 1. Ethiopia would have, therefore, ranked second, thus obtaining a direct qualification. The exclusion of Ethiopia from the control group is, therefore, incorrect.

G.34 Angola, Group C, ACN 2015

$N_C = 4$, $L_C = 2$, $\mathbb{I}_C^V = 1$, $\mathbb{I}_{AGO,2015}^{FP} = 0$; table before the last match day: Burkina Faso$_1$ 10, Angola$_2$ 5, Gabon$_3$ 9, Lesotho$_4$ 2; last match day: Burkina Faso$_1$ vs. Angola$_2$, Gabon$_3$ vs. Lesotho$_4$. If $\boldsymbol{\omega} = (0\ p)'$,

[37] See https://athlet.org/football/can/2012/qualifiers/group-h (last accessed on April 7, 2024).

[38] The same argument applies neither to Benin, which could not be admitted to the virtual playoff group with more than 5 adjusted points, nor to Rwanda, which could not be admitted to the virtual playoff group with more than 3 adjusted points, as less than 6 adjusted points necessarily implied elimination (see Table E.5).

[39] See https://athlet.org/football/can/2012/qualifiers/group-i (last accessed on April 7, 2024).

[40] See https://athlet.org/football/can/2015/qualifiers/group-b (last accessed on April 7, 2024).

Angola would have ranked third, thus gaining access to the virtual playoff group with 8 points; 8 points were potentially sufficient to qualify (see Table E.5). The exclusion of Angola from the control group is, therefore, incorrect.

G.35 Ivory Coast, Group D, ACN 2015

$N_D = 4$, $L_D = 2$, $\mathbb{I}_D^V = 1$, $\mathbb{I}_{CIV,2015}^{FP} = 1$; table before the last match day: Ivory Coast$_1$ 9, Cameroon$_2$ 13, DRC$_3$ 6, Sierra Leone$_4$ 1; last match day: Ivory Coast$_1$ vs. Cameroon$_2$, DRC$_3$ vs. Sierra Leone$_4$. If $\boldsymbol{\omega} = (0\ 3)'$, the DRC and Ivory Coast would have had 9 points each, and the Ivory Coast would have ranked third, as the DRC was dominating the Ivory Coast with respect to the number of away goals in direct confrontations (see Table C.1), as DRC had won 3-4 in the Ivory Coast, and the Ivory Coast had won 1-2 in the DRC.[41] The Ivory Coast would have gained access to the virtual playoff group with 9 points, which were potentially insufficient to qualify (see Table E.5). The exclusion of the Ivory Coast from the treatment group is, therefore, incorrect. The final table of this group was: Cameroon 14 (qualified), Ivory Coast 10 (qualified), DRC 9 (qualified via the virtual playoff group) and Sierra Leone 1 (eliminated).

G.36 Ghana, Group E, ACN 2015

$N_E = 4$, $L_E = 2$, $\mathbb{I}_E^V = 1$, $\mathbb{I}_{GHA,2015}^{FP} = 1$; table before the last match day: Ghana$_1$ 8, Togo$_2$ 6, Guinea$_3$ 7, Uganda$_4$ 7; last match day: Ghana$_1$ vs. Togo$_2$, Guinea$_3$ vs. Uganda$_4$. If $\boldsymbol{\omega} = (0\ 3)'$, then Ghana would have ranked third. Ghana would have gained access to the virtual playoff group with 8 points, which were potentially insufficient to qualify (see Table E.5). The exclusion of Ghana from the treatment group is, therefore, incorrect.

[41] See https://athlet.org/football/can/2015/qualifiers/group-d (last accessed on April 7, 2024).

H Treatment dates

We describe here the treatment date for the country-qualification campaign pairs for which this corresponds to a different week from the one of the last match day of the group in which a team was included in the qualification group stage.

H.1 Eritrea, Group 1, ACN 2000

The last match day of Group 1 was on June 19, 1999; Eritrea was admitted to a playoff group with Senegal and Zimbabwe (with matches from July 4, to August 21, 1999).[42] The treatment status of Eritrea (eliminated) was determined on August 15, 1999, after its defeat against Zimbabwe.

H.2 Senegal, Group 5, ACN 2000

The last match day of Group 5 was on June 19, 1999; Senegal was admitted to a playoff group with Eritrea and Zimbabwe. The treatment status of Senegal (qualified) was determined on August 21, 1999, after its win against Zimbabwe.

H.3 Zimbabwe, fictitious group, ACN 2000

Zimbabwe did not participate in the qualification group stage, but it was directly admitted to a playoff group with Eritrea and Senegal. The treatment status of Zimbabwe (eliminated) was determined on August 21, 1999, after its defeat against Senegal.

H.4 Benin, Mali, Togo and Uganda, ACN 2008

The matches of the last match day of the various groups giving access (for the runner-up) to the virtual playoff group were played between September 7 to 9, 2007, and October 12 and 13, 2007 (for Groups 2 and 9). The first three teams in the virtual playoff group qualified for the final phase of the ACN 2000. Before the last match day, the runner-up of Group 4, either Sudan or Tunisia, was already certain to rank first in the virtual playoff group. A team ranking second or third after the matches played between September 7 and 9, 2007 could still qualify, while a team ranking fourth or more was certainly eliminated. The (incomplete) table of the virtual playoff group was: Tunisia 13, South Africa 11, Uganda 11, Equatorial Guinea 10, DRC 9, Mozambique 9, Togo 9, Eritrea 9, Algeria 8. As Burundi had drawn 1-1 against Egypt in the last match of Day 5 in Group 2[43], the runner-up (Botswana, Burundi, and Mauritania) of Group 2 could not have more than 10 points, hence they could not obtain a qualification via the virtual playoff group. However, both

[42] See https://fr.athlet.org/football/can/2000/qualifications/barrages (last accessed on April 8, 2024).

[43] See https://athlet.org/football/can/2008/qualifiers/group-2 (last accessed on April 8, 2024).

Botswana and Burundi were still in a position to obtain a qualification as the winner of Group 2[44], so that their treatment date was determined only on October 13, 2007, after the last match day of their group. In Group 9, both Benin and Togo could qualify as runner-ups (with 11 and 12 points respectively), while Mali was certain to be eliminated as a runner-up (with 10 points), but it could still win the group. Therefore, South Africa was also qualified, while the treatment status of Uganda was uncertain, and Equatorial Guinea, DRC, Mozambique, Togo, Eritrea, and Algeria were eliminated.

The last match day of Group 9 on October 12, 2007 determined the qualification of Mali as the winner of the group, the qualification of Benin (runner-up with 11 points, but with a better goal difference than Uganda), the elimination of Togo (third with 9 points), and of Uganda.

[44]Botswana would have certainly won the group by defeating Egypt, provided that Burundi did not win its match; if both teams had won on the last match day, they would have tied at 10 points, with a perfect equilibrium in direct confrontations, so that the winner of the group would have been determined by the total goal difference.

I Teams with concrete chances of qualification

We briefly describe here seven qualification-campaign pairs *not* included in the original estimation sample in DC^2, corresponding to teams that unambiguously had concrete chances of qualification or elimination before the last match day. The details about these cases are provided in Section G above.

I.1 Mozambique, Group 1, ACN 2000

Mozambique 2000 failed to qualify, and it is not included by DC^2 in the control group. Mozambique belonged to a group of three teams, where the first team obtained a direct qualification, and the runner-up joined a real playoff group with other two teams. Mozambique ranked second before the last match day, and it would have joined the real playoff group if it had won or drawn on the last match day against Eritrea. Thus, Mozambique would have joined the real playoff group in two of the three combination of points on the last match day of its group.[45] This real playoff group offered one additional slot in the final phase of the ACN 2000. See Section G.4 for more details.

I.2 Burundi, Group 5, ACN 2000

Burundi 2000 failed to qualify, and it is not included by DC^2 in the control group. Burundi belonged to a group of three teams, where the first team obtained a direct qualification, and the runner-up joined a real playoff group with other two teams. Burundi ranked second before the last match day, and it would have joined the real playoff group if it had won or drawn on the last match day against Senegal. Thus, Burundi would have joined the real playoff group in two of the three combination of points on the last match day of its group. This real playoff group offered one additional slot in the final phase of the ACN 2000. See Section G.5 for more details.

I.3 Zimbabwe, fictitious group, ACN 2000

Zimbabwe 2000 failed to qualify, and it is not included by DC^2 in the control group. Zimbabwe was granted access to a real playoff group with two other teams (see Section B.2). This real playoff group offered one additional slot in the final phase of the ACN 2000. See Section G.7 for more details.

I.4 Kenya, Group 3, ACN 2002

Kenya 2002 failed to qualify, and it is not included by DC^2 in the control group. Kenya belonged to a group of four teams, where the first two teams obtained a direct qualification, and the others were eliminated.

[45] Mozambique lost 1-0 against Eritrea, with a goal scored after 81 minutes (see https://en.wikipedia.org/wiki/2000_African_Cup_of_Nations_qualification), which determined its elimination.

Kenya ranked second before the last match day, and it would have qualified in one of the following cases: (*i*) it had won on the last match day, (*ii*) it had drawn on the last match day, and Gabon had either lost or drawn against Morocco. Thus, Kenya would have qualified in five of the nine combination of points on the last match day of its group. See Section G.9 for more details.

I.5 Ethiopia, Group 2, ACN 2004

Ethiopia 2004 failed to qualify, and it is not included by DC^2 in the control group. It belonged to a group of four teams, where the first team obtained a direct qualification, the runner-up joined a virtual playoff group, and the other two teams were eliminated. Ethiopia ranked first (with a tie with Guinea) before the last match day, and it would have qualified in one of the following cases: (*i*) it had won on the last match day, (*ii*) it had drawn on the last match day. Thus, Ethiopia would have certainly qualified in six of the nine combinations of points on the last match day of its group. See Section G.10 for more details.

I.6 Guinea, Group 2, ACN 2004

Guinea 2004 qualified for the final phase of the ACN, and it is not included by DC^2 in the treatment group. It belonged to a group of four teams, where the first team obtained a direct qualification, the runner-up joined a virtual playoff group, and the other two teams were eliminated. Guinea ranked second (with a tie with Ethiopia) before the last match day, and it would have been certainly eliminated if (*i*) it had lost on the last match day, and virtually eliminated if (*ii*) it had drawn on the last match day.[46] Thus, Guinea would have been certainly or almost virtually eliminated in six of the nine combinations of points on the last match day of its group. See Section G.11 for more details.

I.7 Togo, Group 9, ACN 2008

Togo 2008 failed to qualify, and it is not included by DC^2 in the control group. It belonged to a group of four teams, where the first team obtained a direct qualification, the runner-up joined a virtual playoff group with nine other teams offering three additional slots in the final phase of the ACN, and the other two teams were eliminated. Togo was ranked second (with a tie with Mali) before the last match day, and it would have certainly qualified in one of the following cases: (*i*) it had won at home on the last match day. Thus, Togo would have certainly qualified in three of the nine combinations of points on the last match day of its group.[47] See Section G.9 for more details.

[46] In this case, it would have joined the virtual playoff group with 10 points; with these number of points, it would have been eliminated in 97.51 percent of the combination of points of the last match day; in 2.49 percent of the cases, it would have tied with three to six other runner-ups, with only the team with the best goal difference obtaining an indirect qualification; these shares can be computed from Table E.1.

[47] In case of a draw, it would have ranked second with 10 points if Benin drew or lost its last match, thus with possible chances of qualification also via the virtual playoff group (see Table E.2).

J Mistakes in the code for Eq. (2)

J.1 Lags of the dependent variable

DC^2 unnecessarily drop the first four weeks for each country-qualification campaign pairs, e.g., $4 \times 109 = 436$ observations in the second-data column of Table 5, when including four lags of the dependent variable, as the lags are created on a dataset that only includes the country-weeks pairs belonging estimation sample. We adjust the code to include a sufficient number of weeks before the treatment date. This mistake affects Cols. (2), (4), (6) and (8) in Table 5 in DC^2, and all the data columns in Tables 6 to 8.

J.2 Changing names for the same country

The file `CAN_57_15.xls` in the replication package, which is used to define the country-qualification campaign pairs corresponding to a (possible) first or overdue qualification for the final phase of the ACN, which are used in Table 6 in DC^2, inconsistently reports the name of some countries in different years, e.g., DRC and Zaire, Congo and Congo-Brazzaville; this implies, for instance, that DC^2 classify the DRC for the ACN 1998 as an instance of an overdue qualification, even though it had actually qualified for the final phase of the ACN 1996 (and being recorded as Zaire). This mistake affects Cols. (2) and (3) in Table 6 in DC^2.

J.3 Merge many-to-many (`m:m`) in Stata

The second (third) data column in Table 7 in DC^2, "refers to conflict events taking place in locations inhabited by ethnic groups with strong (weak) political power" (DC^2, p. 1597); the code that generates the estimation sample is based on `0-dataset-construction-ACLED.do`, with a merge many-to-many (`m:m`), using the identifier of each ethnic group, between a new event-level dataset (`EPR_ACLED_all.dta`), where a single conflict event can be associated to multiple ethnic groups, and a new ethnic group × period dataset (`EPR_groups.dta`), where the political power of each group is measured at different periods in time from 1946 to 2017 (see https://icr.ethz.ch/data/epr/core/2018.html, last accessed on June 5, 2024), but with these periods that are *not* reported in the dataset `EPR_groups.dta`; as there are, in general, many more conflict events than periods associated to each ethnic group, this merge assigns the political status of last observation for an ethnic group in `EPR_groups.dta` to most conflict events associated to that ethnic group (see Stata, 2015, p. 454, on this), and the merge is unstable if the sorting of the observations in the two datasets is modified when running the code; we have recovered the underlying data from Cederman et al. (2010), and merged the two datasets using the identifier of each ethnic group and the year (more precisely, the political status of an ethnic group is measured in the year before the conflict event). This mistake affects Cols. (2) and (3) in Table 7 in DC^2.

K Qualifications for the FIFA World Cup

DC[2] also define the treatment for the qualification group stage for the FIFA World Cup from 1998 to 2014. The country-qualification campaign pairs in the treatment and the control groups are listed in the notes to Table A.28 in the online Appendix. There are two exclusion errors related to the World Cup 2006, which we detail here.

K.1 DRC, Group 2, World Cup 2006

Groups of six teams, the first qualified for the final phase of the FIFA World Cup.

Table before the last match day: Cape Verde$_1$ 10, Ghana$_2$ 18, South Africa$_3$ 15, DRC$_4$ 15, Uganda$_5$ 7, Burkina Faso$_6$ 12; last match day: Cape Verde$_1$ vs. Ghana$_2$, South Africa$_3$ vs. DRC$_4$, Uganda$_5$ vs. Burkina Faso$_6$.

If $\boldsymbol{\omega} = (3\ 0\ p)'$, then both Ghana$_2$ and DRC$_4$ would have had 18 points. They had both obtained 2 points in direct confrontations (Ghana vs. DRC 0-0 and DRC vs. Ghana 1-1).[48] As FIFA was not using the number of away goals as a tiebreaker in direct confrontations,[49] DRC could have qualified with a better goal difference than Ghana. Before the last match day, Ghana had 13 goals scored and 4 against (goal difference of 9), and the DRC had 12 goals scored and 8 against (goal difference of 4). Therefore, if $\boldsymbol{\omega} = (3\ 0\ p)'$, and the sum of the margin of victory of the DRC and of Cape Verde (against Ghana) was of at least 6 goals, then the DRC would have qualified for the World Cup. The exclusion of DRC from the control group is, therefore, incorrect.

K.2 Ghana, Group 2, World Cup 2006

See Section K.1. The exclusion of Ghana from the treatment group is, therefore, incorrect.

[48] See https://athlet.org/football/world-cup/2006/qualifiers/caf/second-round-group-2 (last accessed on April 19, 2024)

[49] See Article 7 of the Regulations for the FIFA World 2006, available at: https://www.uefa.com/newsfiles/19085.pdf (last accessed on April 19, 2024).

L Estimates and placebos (correct sample and dates)

Table L.1: (Table 5 in DC[2]) ACN Qualification and Conflict
(correct sample and treatment dates)

	Dependent variable							
	log(1+number of events)						Number of events	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-qualification	-0.131	-0.086					-0.329	-0.231
	(0.059)	(0.035)					(0.155)	(0.127)
12 weeks before qualification			0.107	0.075				
			(0.058)	(0.043)				
1-12 weeks post-qualification (a)					-0.127	-0.095		
					(0.057)	(0.034)		
13-25 weeks post-qualification (b)					-0.135	-0.078		
					(0.072)	(0.044)		
Long-run impact	-0.131	-0.147	0.107	0.076				
	(0.059)	(0.060)	(0.058)	(0.043)				
Pr $> F$ H$_0$: $a = b$					0.887	0.652		
Regression method	OLS	OLS	OLS	OLS	OLS	OLS	Neg. binomial	
Country × qualification campaign	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Calendar-month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4 lags of dep. variable	No	Yes	No	Yes	No	Yes	No	Yes
Pr$> F$ H$_0$: 4 lags jointly$= 0$		0.00		0.00		0.00		
Observations	7,000	7,000	3,500	3,500	7,000	7,000	7,000	7,000
Within R^2	0.004	0.087	0.003	0.041	0.002	0.087		
Placebo treatment	0.107	0.075			0.107	0.075	0.226	0.149
	(0.058)	(0.043)			(0.058)	(0.043)	(0.147)	(0.127)
p-value	0.065	0.082			0.065	0.082	0.123	0.240

Notes: see Table 5 in DC[2]; we report, for each data column, the estimated coefficient, standard error and p-value of the placebo treatment obtained from the estimation of Eq. (3); this is not reported for Cols. (3) and (4), which are themselves the placebo tests for Cols. Cols. (1) and (2).

Table L.2: (Table 6 in DC[2]) Overdue and first qualification effects (correct sample and treatment dates)

	Dependent variable				
	log(1+number of events)				
	(1)	(2)	(3)	(4)	(5)
Post-qualification (a)	-0.086	-0.153	-0.086	-0.213	-0.097
	(0.035)	(0.056)	(0.036)	(0.147)	(0.039)
Post-qualification × overdue (b)			-0.146		
			(0.071)		
Post-qualification × first time (c)					-0.162
					(0.170)
Pr > F H_0: a and b (c) jointly= 0			0.026		0.047
Sample	Full	Overdue qualification	Full	First qualification	Full
Country × qualification campaign	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes
Calendar-month FE	Yes	Yes	Yes	Yes	Yes
4 lags of dep. variable	Yes	Yes	Yes	Yes	Yes
Pr> F H_0: 4 lags jointly= 0	0.00	0.00	0.00	0.00	0.00
Observations	7,000	3,700	7,000	1,150	7,000
Within R^2	0.087	0.087	0.073	0.127	0.088
Placebo treatment		-0.007	-0.112	0.078	-0.045
		(0.062)	(0.070)	(0.070)	(0.059)
p-value		0.914	0.111	0.276	0.447

Notes: see Table 6 in DC[2]; we report, for each data column, the estimated coefficient, standard error and p-value of the placebo treatment obtained from the estimation of Eq. (3); for Col. (3), these correspond to the coefficient of the interaction between the placebo treatment and a dummy for an overdue qualification; for Col. (5), these correspond to the coefficient of the interaction between the placebo treatment and a dummy for a country that had never qualified to the final phase of the ACN; we do not report the results of the placebo test for Col. (1), as this simply reproduces Col. (2) in Table L.1.

Table L.3: (Table 7 in DC[2]) Ethnic conflict, ethnic political power, and linguistic diversity (correct sample and treatment dates)

	Dependent variable				
	log(1+number of events)				
	(1)	(2)	(3)	(4)	(5)
Post-qualification	-0.018	-0.034	-0.052	-0.026	-0.055
	(0.012)	(0.013)	(0.024)	(0.018)	(0.028)
Long-run impact	-0.024	-0.066	-0.081	-0.042	-0.092
	(0.018)	(0.035)	(0.039)	(0.033)	(0.048)
Event definition	Ethnic	Strong political power	Weak political power	No linguistic diversity	High linguistic diversity
Country × qualification campaign	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes
Calendar-month FE	Yes	Yes	Yes	Yes	Yes
4 lags of dep. variable	Yes	Yes	Yes	Yes	Yes
Pr> F H_0: 4 lags jointly= 0	0.00	0.00	0.00	0.00	0.00
Observations	7,000	7,000	7,000	7,000	7,000
Within R^2	0.055	0.103	0.065	0.066	0.086
Placebo treatment	-0.017	0.023	0.057	0.014	0.079
	(0.023)	(0.016)	(0.036)	(0.014)	(0.042)
p-value	0.469	0.145	0.113	0.320	0.062

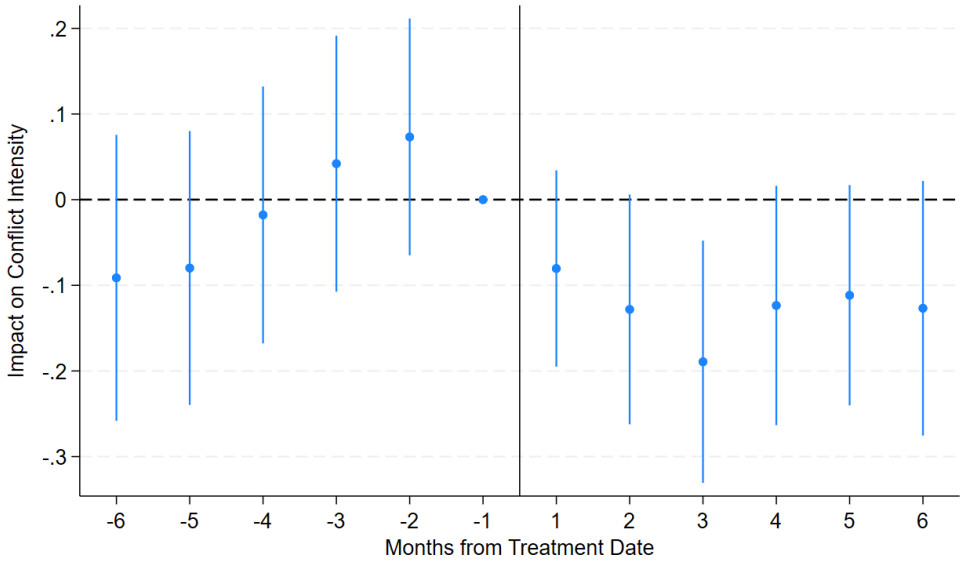
Notes: see Table 7 in DC[2]; we report, for each data column, the estimated coefficient, standard error and p-value of the placebo treatment obtained from the estimation of Eq. (3).

Table L.4: (Table 8 in DC[2]) Ethnic conflict, ethnic political power, and linguistic diversity (correct sample and treatment dates)

	Dependent variable				
	log(1+number of events)				
	(1)	(2)	(3)	(4)	(5)
Post-qualification	-0.085	-0.082	-0.034	-0.026	-0.011
	(0.035)	(0.034)	(0.017)	(0.014)	(0.010)
Long-run impact	-0.144	-0.139	-0.048	-0.036	-0.016
	(0.061)	(0.061)	(0.025)	(0.019)	(0.013)
Omitted observations	ACN weeks	None	None	None	None
Model specification	Baseline	Treatment interacted ACN weeks	Baseline	Baseline	Baseline
Fatality threshold	None	None	> 10	> 25	> 50
Country × qualification campaign	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes
Calendar-month FE	Yes	Yes	Yes	Yes	Yes
4 lags of dep. variable	Yes	Yes	Yes	Yes	Yes
Pr$> F$ H_0: 4 lags jointly$= 0$	0.00	0.00	0.00	0.00	0.00
Observations	6,627	7,000	7,000	7,000	7,000
Within R^2	0.086	0.087	0.042	0.039	0.039
Placebo treatment	0.075	0.075	-0.015	-0.008	-0.006
	(0.042)	(0.042)	(0.023)	(0.019)	(0.011)
p-value	0.082	0.082	0.510	0.678	0.581

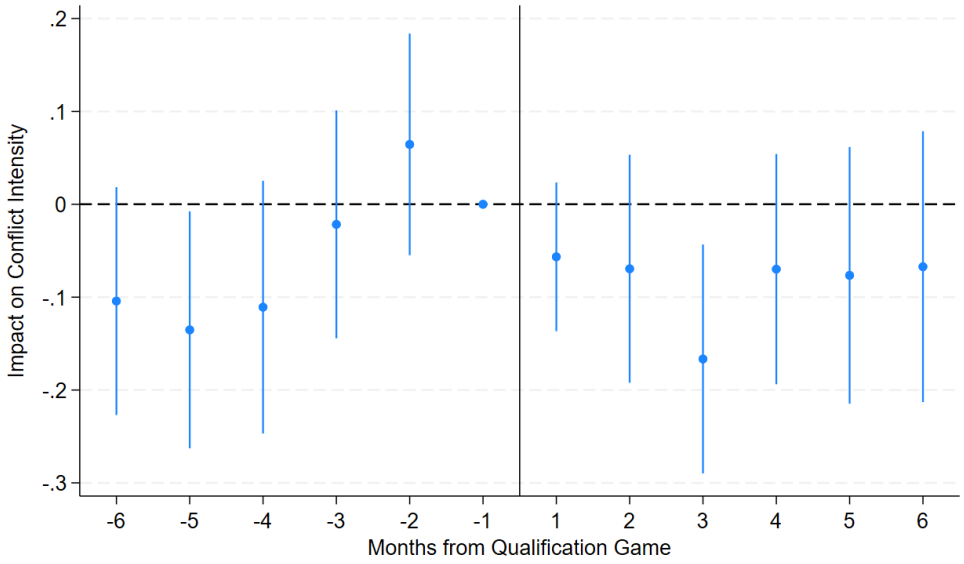
Notes: see Table 8 in DC[2]; we report, for each data column, the estimated coefficient, standard error and p-value of the placebo treatment obtained from the estimation of Eq. (3).

Figure L.1: (Panel A of Figure 4 in DC[2])
Number of conflict events before and after qualification
Treated and control countries (correct sample and treatment dates)



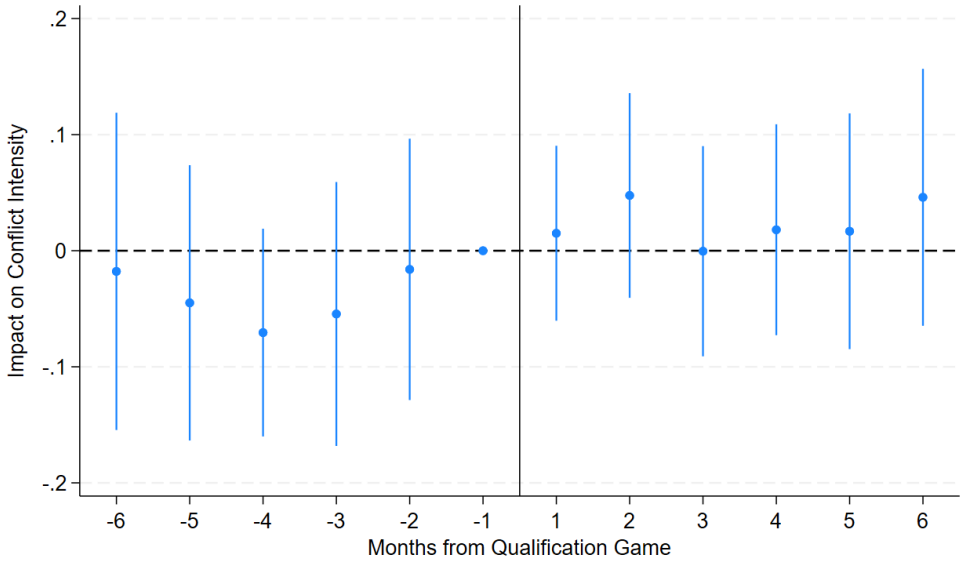
Notes: see Figure 4 in DC[2]

Figure L.2: (Panel B of Figure 4 in DC[2])
Number of conflict events before and after qualification
Treated countries only (correct sample and treatment dates)



Notes: see Figure 4 in DC[2]

Figure L.3: (Panel C of Figure 4 in DC[2])
Number of conflict events before and after qualification
Control countries only (correct sample and treatment dates)



Notes: see Figure 4 in DC[2]

M Corrigendum (Depetris-Chauvin et al., 2024)

M.1 Relationship between corrigendum and original article

The algorithm does not introduce (C.1) for teams admitted to a virtual playoff group, but this condition can be violated even when condition (C2.b) is met, as in the case of Cameroon 2012.[50] This introduces a clear inconsistency with the descriptions of the sample selection criteria in DC^2, which always refer to the indeterminacy of the qualification status before the last match day. The use of conditions (C.2a) and (C.2b), which are based on the points in the final table of a group or of the virtual playoff group, raises four main issues related to their consistency with the descriptions of the estimation sample included in the original article: (*i*) the algorithm is dependent on the situation after the last match day,[51] while the original article clearly emphasized that the estimation sample was based on an analysis of the situation before the last match day; (*ii*) the algorithm does not restrict the sample to country-qualification campaign pairs that had "concrete chances of qualifying" (DC^2, p. 1591), e.g., a team that needed to win the last match by a margin of at least 17 goals to qualify can satisfy both (C.1) and (C.2a);[52] (*iii*) the algorithm excludes country-qualification campaign pairs that clearly had concrete chances of qualifying before the last match, but that violate (C.2a) or (C.2b), as the Authors themselves acknowledge; notably, the corrigendum states that Kenya 2012 and Togo 2015, which were included in the control group of the estimation sample in DC^2 even though they violate (C.2a), as "arguably constituting close cases" (p. 3). To us, this demonstrates that the use of (C.2a) is inconsistent with the sample selection criteria described in the original article; (*iv*) an algorithm relying on the points in the final table arbitrarily fails to treat consistently country-qualification campaign pairs that were in an identical situation before the last match day, as described in Section 3 above.

M.2 Sample generated by the algorithm in the corrigendum

The Authors acknowledge 12 instances (out of 109 country-qualification campaign pairs) in which the original estimation sample in DC^2 and the estimation sample used in the corrigendum diverge. Notably, at p. 3-4, Depetris-Chauvin et al. (2024) describe six inclusion errors,[53] which "are balanced out" (p. 4) by six exclusion errors. Importantly, an application of the algorithm described in the corrigendum does *not* generate the sample used in the corrigendum. In particular, Sudan 2004 (see Section G.12), Ethiopia 2004 (see Section

[50] Similarly, the converse can also happen: Congo 2004 satisfies (C.1), as demonstrated in Section G.14, but it violates (C2.b), as it had 4 points less than Zimbabwe, the only team that qualified via the virtual playoff.

[51] The results of the last match day can also determine whether the second or the third part of the algorithm applies, and this is crucial as the second part imposes (C.1), while the third part does not; hence, whether (C.1) applies or not depends on the results of the last match day.

[52] We are referring here to Madagascar 2004 (see Section A.2.2).

[53] Four are related to a violation of (C.1), and correspond to four of the five cases described in Section F, and two to a violation of (C.2a); an inclusion error (Cameroon 2012) refers to a violation of (C.1) in a part of the algorithm that does *not* include this condition.

G.10), Eswatini 2004 (see Section G.16), Tanzania 2008 (see Section G.20) and Togo 2008 (see Section G.22) are not included in the sample used in the corrigendum, but they all respect the two conditions described in the algorithm, as they had (*i*) an undetermined qualification status, and they also respect the condition (*ii*) related to the points in the final table. Their exclusion from the sample used the corrigendum seems inconsistent with the algorithm described in the corrigendum itself. Additional inconsistencies relate to two country-qualification campaign pairs that are included both in the original estimation sample, and in the sample used in the corrigendum. In particular, the inclusion of Mali 2004 (treated) and Uganda 2004 (control) in both DC^2 and Depetris-Chauvin et al. (2024) reveals these additional inconsistencies.

Mali 2004 was included in Group 6 with three other teams; the final table of its group was: Mali 13 (qualified), Zimbabwe 13 (qualified via the virtual playoff group), Seychelles 6 (eliminated), Eritrea 3 (eliminated).[54] Thus, Mali ended up 7 points ahead of Seychelles, "the team in the group that had the most points while still failing to qualify" (Depetris-Chauvin et al., 2024, p. 2), so it should have been excluded from the treatment group as it violates (C.2a). The only way to justify the inclusion of Mali 2004 in sample is to interpret the algorithm described in the corrigendum should be interpreted as considering only teams that failed to qualify *directly*, i.e., we should consider the difference in points between Mali and Zimbabwe, which stands at zero. However, this interpretation would imply that four additional country-qualification campaign pairs have been incorrectly excluded from the treatment group of the sample: Zambia 2008 (see Section G.25), Ghana 2012 (see Section G.32), Zambia 2012 (see Section G.29) and Ivory Coast 2015 (see Section G.35). These four cases correspond to country-qualification campaign pairs that meet (C.1), and which had, in the final table, no more than three points above a team which qualified via the virtual playoff group, thus satisfying (C.2a) according to the interpretation that justifies the inclusion of Mali 2004 in the sample. Thus, we have identified either an additional inclusion error, or four additional exclusion errors.

The case of Uganda 2004 has already been discussed in Section A.1: its inclusion in the estimation sample can be justified only if (C.1) is assessed before its own last match, and not before the last match day. However, the inclusion of Uganda 2004 in the estimation sample implies that two more country-qualification campaign pairs have been incorrectly excluded from the treatment group: Ghana 1998 (see Section A.1.2), Nigeria 2004 (see Section A.1.6). Both cases satisfy (C.1), interpreted in the way that justifies the inclusion of Uganda 2004 in the sample, and also meet (C.2a). Thus, we have identified either an additional inclusion error, or two additional exclusion errors.

A consistent application of the algorithm reveals up to two exclusion errors and between six and 12 inclusion errors in in the sample used in the corrigendum. In other words, a consistent application of the algorithm in Depetris-Chauvin et al. (2024) generates a sample that diverges from the one used to generate the estimate presented in the corrigendum for eight to 12 country-qualification campaign pairs. If we compare the sample generated by a consistent a application of the algorithm with the original estimation sample in DC^2, the two diverge for 20 to 24 country-qualification campaign pairs, i.e., 18.3 to 20.0 percent of the original estimation sample. Thus, the corrigendum does not, in our view, clarify how the sample in the original article was built, but it describes an algorithm that does not generate either the sample used in DC^2, nor the one used in the estimates presented in the corrigendum itself. The estimates presented in

[54] See https://athlet.org/football/can/2004/qualifiers/group-6 (last accessed on Dec. 6, 2024).

Depetris-Chauvin et al. (2024) are uninformative about those that can be obtained when correctly applying the sample selection criteria described in DC^2, and even about the results obtained from the sample obtained when applying the algorithm itself.

M.3 Missing revised replication package

The "revised replication package" (p. 5), which should underpin the corrigendum, is not available online. It is, therefore, impossible to understand how, and using which raw data, the Authors have dealt with the most difficult part of the algorithm, which is the identification of the teams with an undetermined qualification status before the last match day. This requires a substantial effort of data collection and coding, that we have undertaken for our replication, and that is not documented in the corrigendum. The unavailability of the replication package also prevents us from understanding why some country-qualification campaign pairs have been classified in an inconsistent way, as the cases of Mali 2004 and Uganda 2004 described above reveal.

The corrigendum repeatedly explains the exclusion of some country-qualification campaign pairs from the original estimation sample by referring to criteria, e.g., the goal difference or the ranking in the final table of a group, that are *not* considered in the algorithm that the Authors have described. For instance, the corrigendum reads:

> "Namibia and Kenya, in 2002, finished within three and two points of the last qualified team, respectively. They *were overlooked in all likelihood because they finished in fourth place, and with negative goal differentials* they would have had to reverse." (Depetris-Chauvin et al., 2024, p. 4, emphasis added).

In our view, factors that are not used by the algorithm described in the corrigendum cannot be used to explain instances in which the algorithm incurs into exclusion errors. To us, this further demonstrates the importance of a consistent application of a clear definition of the sample selection criteria, based on the original article. This is precisely what our paper does, on the basis of a code that is publicly available, and that starts from the raw data that are necessary to determine whether a team had an undetermined qualification status before the last match day.

M.4 Concordance between the algorithm and our code

Our analysis uncovers five inclusion and 36 exclusion errors in the original estimation sample in DC^2 (see Section 3). It is interesting to relate them with the sample that is obtained with a consistent application of the algorithm described in the corrigendum. As our definition relies only on (C.1), four of the five inclusion errors that we uncovered also arise with a consistent application of the algorithm.[55] Out of the 36 cases

[55]We consider here that Uganda 2004 represents an inclusion error; the only exception is represented by Cameroon 2012, which violates (C.1), but was admitted to a virtual playoff group, where it respects (C2.b); the corrigendum describes as an exclusion error in DC^2, but it actually satisfies the inclusion criteria set by the algorithm, as these depend only on (C2.b) for teams admitted to a virtual playoff group.

representing exclusion errors in the original sample, 12 cases unambiguously satisfy also (C.2a)-(C.2b), so they belong to the sample generated by the algorithm, once this is consistently applied, and four more cases also satisfy (C.2a) if Mali 2004, which belongs to the sample used in the corrigendum, is not an inclusion error, as discussed above. Thus, 16 out of 36 exclusion errors also belong to the sample that is generated by a consistent application of the algorithm. Five more cases relate to the qualification campaign for the ACN 2000, the only one to include a real playoff group. The algorithm described in the corrigendum confirms that the sample in the original article has been built without accounting for the specific format of this edition.[56]

Thus, a consistent application of the algorithm described in the corrigendum also identifies four of the five inclusion errors and 21 out of 36 exclusion errors that we uncover in the original estimation sample.[57] Thus, the two approaches coincide for 25 out of 41 cases that we have described in Section 3 above. In other words, a substantial portion of the discordance between the sample that is obtained by our code, which only uses (C.1), and the sample used in the original article arises from an inconsistent application of the algorithm described in the corrigendum.

The remaining 15 country-qualification campaign pairs that constitute exclusion errors according to our definition, which is based on the descriptions of the sample selection criteria described in the original article, as they satisfy (C.1), but they do not respect (C.2a). None of these 15 country-qualification campaign cases had chances of qualification before the last match day that were lower than those of Madagascar 2004 (control), or chances of elimination lower than those of Ivory Coast 2008 (treated), which are included in the original estimation sample in DC^2, coherently with the algorithm described in the corrigendum. Madagascar 2004 needed to win by a margin of at least 17 goals to qualify (see Section A.2.2), and Ivory Coast would have been eliminated by losing by a margin of at least six goals against Gabon, after having won 5-0 in the first leg (see Section A.2.4). This, again, confirms that the refinement introduced by (C.2a) is not related to the chances of qualification or elimination before the last match day, and it is, thus, inconsistent with the descriptions of the sample selection criteria provided in the original article by the Authors themselves.

[56]Notice that all these five cases actually satisfy (C.2a), once this is interpreted in the same way that justifies the inclusion of Mali 2004 in the sample, e.g., Mozambique 2000 ended up just one point below Eritrea 2000, which had gained admission to the real playoff group (see Section G.4).

[57]Cameroon 2012 is presented an inclusion error in the corrigendum, but our reading of the algorithm implies that its inclusion in the original estimation sample is consistent with the algorithm, even though its qualification status was fully determined before the last match day.