# Can Soccer Predictions Boost the Use of Indirect Comparison to Assess the Safety Of Drugs?

M. Soledad Cepeda*[,1], Daniel Fife[1], Victor Lobanov[1] and Alex Sutton[2]

[1]*Johnson & Johnson Pharmaceutical Research and Development, 1125 Trenton Harbourton Rd, Titusville, NJ 08560, USA;* [2]*Department of Health Sciences, University of Leicester, 2nd Floor (Room 214e), Adrian Building, University Road, Leicester LE1 7RH, UK*

**Abstract:** Randomized comparisons to establish the relative safety of competing interventions are scarce. An approach to estimate relative treatment effects in the absence of head-to-head trials is to utilize indirect comparisons and extended meta-analytic techniques, but these techniques are underused.

We provide a non-technical account using analogies to predicting results in sports matches, which we hope makes these methods more accessible. The ideas in this example may be familiar to readers who may have considered an indirect comparison when trying to predict the outcome of a sports match. We used the first round of the soccer world cups of 2006 and 2010. In addition, we provide a simplified clinical example.

We found that even in circumstances in which the signal, i.e. strength of the team or the relatively safety of the medication, can easily be diluted or even lost, there is a fair agreement between the predictions of the indirect comparisons and the actual results.

We hope that this example will increase the acceptability of indirect comparisons and extended methods and widen the basis of health care decisions to a broader range of evidence instead of just the evidence from direct placebo-based comparisons.

## INTRODUCTION

Randomized comparisons to establish the relative safety or tolerability of many competing interventions are scarce. In the presence of such a paucity of evidence, an approach to estimating relative treatment effects in the absence of head-to-head trials is to utilize an "adjusted" indirect comparisons approach. However, these methods are underused and difficult to understand. Formal statistical descriptions of this approach have been given elsewhere [1] but below we provide a non-technical account using analogies to predicting results in sports matches, which we hope makes the ideas accessible to those who may feel intimidated by technical descriptions. In addition, we illustrate how an indirect comparison is performed using a clinical example.

Many readers may have considered an indirect comparison when trying to predict the outcome in an event such as a soccer tournament and may have been among the approx 30,000,000 people who watched the 2010 Soccer World Cup, which we will now revisit. Let's assume that we want to predict the last game of group E – a match between the Netherlands and Cameroon. What we do instinctively is find a common comparator, e.g., a team that both the Netherlands and Cameroon have competed against recently, and compare

the outcomes. Since both Cameroon and the Netherlands have played against Denmark, we can predict which team would likely to be a winner based on the scores of their games against Denmark (Fig. **1**). Cameroon lost to Denmark and Denmark lost to the Netherlands, so we predict the Netherlands will win and, in this case our prediction was correct. The Netherlands beat Cameroon 2 to1.

Many factors can compromise validity of indirect comparisons of health interventions. First, an indirect comparison must respect the randomization to ensure balance in patient characteristics across groups. Comparing results of similar treatment arms from different clinical trials is strongly discouraged and can produce biased and overly precise estimates [2] since this does not respect randomization. Hence, comparisons should be based on treatment differences within each randomized trial. Second, the results of trials used in an indirect comparison should be homogeneous. In particular, all factors that affect the response to a treatment, i.e. effect modifiers, must have a similar distribution across the entire set of trials that are being compared. This implies that all considered trials should have similar study designs, populations and outcomes [2, 3].

In the above soccer prediction, we based the analysis on the scores of each of the matches, which is analogous to respecting the randomization in a clinical trial setting (i.e., we did not count all victories by Cameroon and all victories by the Netherlands). Since the Netherlands played Denmark only once and similarly, Cameroon played Denmark only

*Address correspondence to this author at the Johnson & Johnson Pharmaceutical Research and Development, 1125 Trenton Harbourton Rd, Titusville, NJ 08560, USA; Tel: 609 730 2413; Ext: 30001; Fax: 609 730 7927; E-mail: scepeda@its.jnj.com
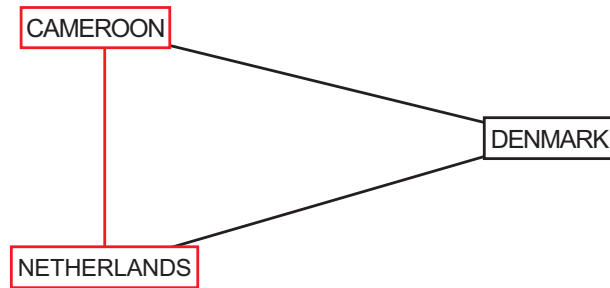
**Fig. (1).** To predict the match between the Netherlands and Cameron we can use the scores of the previous games between Cameroon and Denmark and Netherlands and Denmark.

once, we cannot evaluate the homogeneity of the results (which we can do if there are multiple trials of the same comparisons). However, both the Netherlands and Cameroon played Denmark under similar circumstances. The games were very close in calendar time, weather was similar, and although the matches were played in different cities, none of the teams had a home advantage. These are the types of assessments that are required to address the validity of the indirect comparisons [2].

Indirect comparisons in health care are not performed commonly despite all of their potential benefits, as systematic reviews and meta-analyses often focus on trials that compare one intervention with placebo. We hypothesized that if we could show that indirect comparisons were useful in predicting soccer winners, despite all the difficulties involved in such predictions, it might popularize the technique and increase acceptability and use of indirect comparisons and similar extended methodologies. With such an objective in mind we embark on a further exploration of the analogy.

**MATERIALS AND METHODOLOGY**

We decided to use the first round results from the soccer world cups of 2006 and 2010, which are readily available from the official FIFA website [4]. We restricted our attention to the first round because the rules of the competition were suitable for this exercise. In the first round, teams are divided into groups of four teams, each team in a group plays once against each other team in that group, and ties are permitted. Of the six games played in each group, we predicted the last two games (Fig. **2**). Since the earlier games can provide more than one common comparator, two predictions can be made about the same match. For example, the prediction of the Netherlands and Cameroon match can be based on how both teams fared against Japan or against Denmark. We elected to combine the two predications by averaging them.

This is an illustration of how we did the comparisons. To obtain the first prediction, we considered that the Netherlands beat Japan by one goal and Cameroon lost to Japan by one goal. Based on this indirect comparison, we assigned a score of 1- (-1) = 2 to the match between the Netherlands and Cameroon. For the second prediction, we considered that the Netherlands beat Denmark by two goals and Cameroon lost to Denmark by one goal. Hence, we assigned to the match between the Netherlands and Cameron a score of 2 - (-1) =3. After combining the two predictions by averaging, we predicted Netherlands will win.

**RESULTS**

Overall, we made 32 predictions in a similar fashion and found that the agreement between the predictions and the actual results was 59.4% while the expected agreement by chance was 37.3% (i.e., assuming no correlation between results). This result gives us a kappa of 0.35 and a p-value of 0.0025 (Table **1**), which suggests a fair agreement between our predictions and the actual results.

**Clinical Example**

Let's assume that we want to compare the tolerability of gabapentin, an anticonvulsant, and nortriptyline, an antidepressant. These two medications are commonly used for treatment of neuropathic pain. We arbitrarily selected two randomized parallel studies for this comparison. The first study evaluated gabapentin and placebo [5]; the second one evaluated nortriptyline and placebo [6]. The number of subjects in each treatment arm and the number of subjects who withdrew from the studies due to adverse events can be seen in Table **2**.

To assess the tolerability, we calculated the risk of withdrawing from the studies due to adverse events in each arm and then the risk difference in each trial. To determine how gabapentin compares with nortriptyline, all that is needed is to subtract the risks from the two placebo trials (Table **3**). This indirect estimate suggests that the risk of discontinuing medication because of the adverse events is 1.1% lower for nortriptyline than for gabapentin, but the difference is not statistically significant.

In the clinical setting in addition to the placebo control trials there could be head-to-head trials. In this circumstance one can assess the agreement between the direct and indirect estimates and even combined the indirect and indirect estimates. We found a randomized trial that directly compared the tolerability of nortriptyline and gabapentin [7]. The doses of nortriptyline and gabapentin were similar to the doses used in the placebo controlled trials and the duration of the follow up was similar as well as. This head-to-head study found that the risk of discontinuing medication because of adverse events was 2.6% lower with gabapentin, but it was
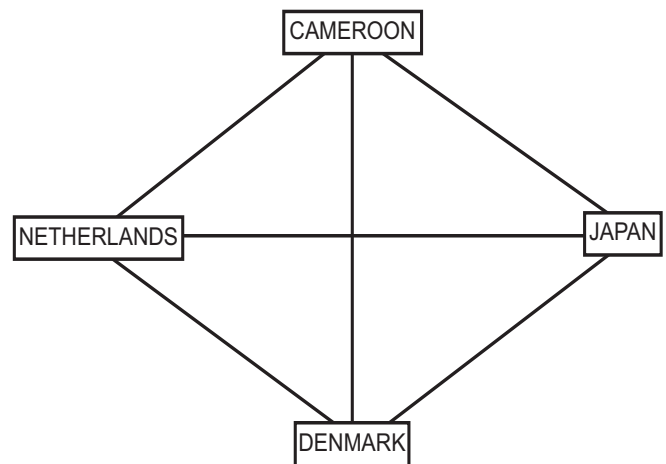


**Fig. (2).** Network of teams and matches in Group E of the 2010 World Soccer Cup. The line connecting the teams represents the games played.

**Table 1.**    **Agreement Among Indirect Comparison Predictions and Observed Match Results**

| | Predicted Results | | | |
|---|---|---|---|---|
| **Observed Results** | **Tie** | **Team in First Position Wins** | **Team in Second Position Wins** | **Total** |
| Tie | 2 | 3 | 3 | 8 |
| Team in first position wins | 1 | 6 | 3 | 10 |
| Team in second position wins | 0 | 3 | 11 | 14 |
| Total | 3 | 12 | 17 | 32 |

Teams were assigned arbitrarily to first or second position for convenience of designation in the above table. In the game between Cameroon and Netherlands described in the text. Cameron is in first position and Netherlands in the second position, a win was predicted for the team in the second position, and it was observed. If there had been perfect agreement between predicted and observed results, all the off-diagonal bolded counts would have been zero.

**Table 2.**    **Description of Direct Comparison Studies Included in Clinical Example and Risk of Withdrawing Due to Adverse Events**

| Study | Treatment | Number of Subjects who Withdrew Due to Adverse events | Number of Subjects in Treatment Arm | Risk of Withdrawal Due to Adverse Events | Risk Difference (95% Confidence Interval) |
|---|---|---|---|---|---|
| 1 [5] | Gabapentin | 15 | 115 | 13.00% | -6.7% (-14% to 0.9%) |
| 1 [5] | Placebo | 7 | 111 | 6.31% | |
| 2 [6] | Nortriptyline | 5 | 38 | 13.16% | -5.66% (-19% to 7%) |
| 2 [6] | Placebo | 3 | 40 | 7.50% | |
| 3 [7] | Nortriptyline | 1 | 38 | 2.63% | 2.63% (-2.4% to 7%) |
| 3 [7] | Gabapentin | 0 | 38 | 0.00% | |

not statistically significant, Table **2**.

The direction of the indirect estimate is in opposite direction to the estimate of the head-to-head trial, but the difference between these two estimates is small (3.7%) and not statistically significant (Table **3**). Therefore, we consider both the direct and the indirect estimates to be in agreement.

It is crucial to emphasize that a proper evaluation of the safety of two treatments should be based on a systematic search of the literature and not on an arbitrary selection of studies. The purpose of this exercise is only to illustrate how the indirect comparison is performed and not to conclude that one medication has a better tolerability profile than another.

## DISCUSSION

Predicting individual soccer results is notoriously difficult [8, 9]. Although, the strength and skills of the team undoubtedly play a role, other variables such as the play of chance, team spirit, personal problems of the players and arbitration decisions (referees are inconsistent and could be biased) have an important role and are difficult to measure or to predict [10]. The low scores typically associated with a soccer match mean the play of chance has a more pro-

nounced effect than in other, high-scoring, sports such as basketball [11]. Therefore, it is not surprising that the observed agreement in predictions of such capricious outcomes as world cup soccer, was lower than the agreement between direct and indirect evidence in the health field, which has been found to be good or very good [12-14].

You pay a price when you perform indirect comparisons, indirect estimates are less precise than direct estimates [1]. This can be appreciated from Table **3**, in which the confidence interval of the indirect comparison estimate is much wider than the confidence interval of the head-to-head estimate.

Often, in the context of comparing health interventions, more than a pairwise comparison is needed. For example, suppose you want to compare the safety profile of all the medications for the treatment of neuropathic pain on the market against each other. In this case, extended meta-analytic methods are used. These methods include mixed treatment comparisons, network meta-analysis and a multiple treatment meta-analysis.

These techniques permit performing simultaneous comparisons of all treatments in the network and often use a bayesian framework, not because bayesian input (prior informa-

**Table 3.**    **Direct and Indirect Comparison Results**

| Direct Comparison Risk (95% Confidence Interval) | Indirect Comparison Risk (95% Confidence interval) | Difference Between Direct and Indirect Estimates (Standard Error) | Z |
|---|---|---|---|
| 2.63% (-2.4% to 7%) | - 6.7% - -5.6% = -1.1% (-13% to 16%) | 2.63% - - 1.1% = 3.73% (0.08) | 0.46 (p=0.64) |

Z= difference between the direct and indirect estimates divided by the standard error of the difference [20].

tion) is needed, (in fact the priors are specified as non–informative), but because of the flexibility of the software used to perform bayesian analyses. The multiple-treatment comparison techniques can deal with trials that have multiple arms by accounting for the correlation due to those multiple arms; they can combine direct and indirect evidence and assess the inconsistency - the disagreement between the direct and indirect evidence; and they can also provide a probabilistic ranking of treatments [15-17]. The threats to the validity of the estimates in this scenario apply to each one of the indirect estimates, so the more complex the network, the more room there is for biased predictions. Efforts to facilitate access to these methods have been made by creating software that not only is friendlier, but that also allows a more efficient and transparent way to check the results of such analyses [18].

## CONCLUSION

We illustrated the method of adjusted indirect comparisons by applying it to a topic of interest to many readers, namely the prediction of soccer match outcomes and to a clinical example. We found that even in circumstances in which the signal, i.e. strength of the team, can easily be diluted or even lost, there is an agreement between the predictions of the indirect comparisons and the actual results. We hope that this example will increase the acceptability of indirect comparisons and extended methodologies and widen the basis of health care decisions to a broader range of evidence instead of just the evidence from direct placebo-based comparisons.

For those who may be inclined to interpret this paper as suggesting a betting system rather than simply giving an accessible illustration of a methodology relevant to comparisons of health interventions, we suggest reading some additional references before betting [19].

## AUTHOR'S CONTRIBUTORS

MSC, DF, and VL were responsible for study concept, study design, data collection, data interpretation, and preparation of the manuscript.

AS was responsible for data analysis, data interpretation, and preparation of the manuscript.

## FUNDING

## REFERENCES

[1]   Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50: 683-91.

[2]   Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 338: b1147.

[3]   Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. Arch Intern Med 2006; 166: 1269-75.

[4]   FIFA. http://www fifa com/ 2010;Available from: URL: http://www.fifa.com/

[5]   Rice AS, Maton S. Gabapentin in postherpetic neuralgia: a randomised, double blind, placebo controlled study. Pain 2001; 94: 215-24.

[6]   Atkinson JH, Slater MA, Williams RA, *et al.* A placebo-controlled randomized clinical trial of nortriptyline for chronic low back pain. Pain 1998; 76: 287-96.

[7]   Chandra K, Shafiq N, Pandhi P, Gupta S, Malhotra S. Gabapentin *vs.* nortriptyline in post-herpetic neuralgia patients: a randomized, double-blind clinical trial--the GONIP Trial. Int J Clin Pharmacol Ther 2006; 44: 358-63.

[8]   Forrest D, Simmons R. Read all about it? Significance 2006; 20-1. doi:10.1111/j.1740-9713.2006.00149.x

[9]   Spiegelhalter D, Ng Y. One match to go. Significance 2009; 151-3. doi:10.1111/j.1740-9713.2009.00387.x

[10]   Goddard J. Is the ref blind? Crime and punishment in English premier league football. Significance 2007; 63-6. doi:10.1111/j.1740-9713.2007.00.227.x

[11]   Heuer C, Muiler C, Rubner O. Soccer: Is scoring goals a predictable Poissonian process? Europhysics Letters 2010; 89: 38007.

[12]   Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003; 326: 472.

[13]   Vandermeer BW, Buscemi N, Liang Y, Witmans M. Comparison of meta-analytic results of indirect, direct, and combined comparisons of drugs for chronic insomnia in adults: a case study. Med Care 2007; 45: S166-72.

[14]   Gartlehner G, Moore CG. Direct *vs.* indirect comparisons: a summary of the evidence. Int J Technol Assess Health Care 2008; 24: 170-7.

[15]   Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29: 932-44.

[16]   Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23: 3105-24.

[17]   Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med 2009; 28: 1861-81.

[18]   Bujkiewicz S, Jones HE, Lai MCW, *et al.* Development of a transparent interactive decision interrogator to facilitate the decision making process in health care. Value Health 2011; In press.

[19]   Spann M, Skiera B. Sports forecasting: "A comparison of the forecast accuracy of prediction markets, betting odds and tipsters". J Forecast 2009; 28: 55-72.

[20]   Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. J Clin Epidemiol 2008; 61: 455-63.