ABSTRACT

Empirical research on the effect of M&A transactions on companies’ performance has not shown clear results of success. It is often assumed that these transactions destroy rather than create value. This study employs meta-analytical techniques to evaluate the outcomes of M&A transactions empirically. This method allows a large quantity of transactions to be examined. Additional factors influencing the performance of M&A transactions are found using a moderator analysis. In total, 55,399 transactions between 1950 and 2010, extracted from 33 previous M&A studies, have been examined. The results of this study confirm findings from previous empirical studies, stating that M&A transactions predominantly do not have a positive impact on the success of a company. A moderator analysis indicates that the type of M&A and the time frame used for measurement influence the success of M&A transactions.

JEL CLASSIFICATION & KEYWORDS
- M10 ■ M16 ■ M20 ■ MERGERS AND ACQUISITION
- META-ANALYSIS ■ PERFORMANCE MEASUREMENT ■ SUCCESS OF M&A

INTRODUCTION

The field of mergers and acquisitions (M&A) has experienced a significant boom due to globalisation. Markets were liberalized and deregulated, trade restrictions dismantled and entry barriers lowered. These developments led to an intensification of competition within the markets, meaning companies now face fiercer competition over prices. A direct consequence of this is consolidation pressure, which leads to an increase in M&A activities (Meckl, 2001; Wirtz, 2014).

As M&As represent one of the most important strategic options for company growth (Denison & Wamser, 2007), it is surprising that the probability of a successful M&A transaction is fairly low (Meckl, Sodeik, & Fischer, 2006). Managers making M&A decisions must be aware that M&As can be risky due to the fact that a variety of variables influence the success of an M&A deal. Examples of such variables are strategic rationale, a consistent project structure, realistic evaluation of the target and effective risk management.

Taking the many influencing variables and the resulting high risk structure of an M&A deal into account, it is not surprising that many empirical studies conclude that M&A transactions destruct value rather than increase it (Harding et al., 2013; Bruner, 2002). Most of these studies attest that more than 50 percent of all M&A transactions were not successful (Weber, Tarba, & Öberg, 2014; Terpitz, 2012). Amongst these are studies which clearly state that the M&A success rate is even lower. Christensen et al. (2011), for instance, claim that the success rate for M&A transactions lies between 10 and 30 percent. It is not surprising that the low success rates of M&As attract serious attention in economic literature. In financial management research, whether or not company transactions are successful is one of the most (common) questions (Brunner, Persteiner, & Wagner, 2014). There are a variety of empirical studies which measure the success or failure of M&A transactions. These studies can be categorized by the grade of complexity and the different methods used to measure performance. These methods can be derived from different theories and can also have an individual focus on particular industries and markets (Hinsen, 2012).

Individual empirical studies may prove or rebut specific hypotheses relating to M&A success. Regarding the high number of studies focused on the success M&As, it makes sense to attempt to increase the level of knowledge in this field by intelligently combining the research already carried out. Consequently, this study aims to contribute to the debate on the success of M&A transactions and to answer the question of whether or not these transactions tend to be successful. This will be done by employing a meta-analysis, which allows as many samples as possible to be taken into account. According to this paper, M&As are considered successful if over 50 percent of the transactions observed are create value. To measure the performance of M&A transactions as a whole, this meta-analysis will be based on a one-variable relationship as opposed to a two-variable relationship, as are most meta-analyses. This method is similar to the one used by Flickinger (2009). The sample includes a total of 55,399 M&A transactions worldwide, taking place between 1950 and 2006. The overall findings indicate that actual M&A transactions tend to be unsuccessful, while additional analyses discover differences in the kind of transaction and the time frame of measurement.

The remainder of the paper is organized as follows: Section 2 gives a brief review of literature that deals with measuring the success of M&As. Section 3 discusses the method and research design used within this paper as well as explaining meta-analysis. Section 4 provides a summary of the results and analyses their limitations and implications.

Literature review

There are many empirical studies concerning the performance of M&A transactions. They differ greatly, however, in regards to methodology and outcomes. There are four main indicators used to measure success: market measures, accounting measures, expert surveys and measures of divestments. The following section presents an overview of findings obtained using these approaches.

When researching M&A success, the most common method is a market-oriented measure called event studies (Glaum & Hutzschenreuter, 2010; Buckmann, 2012). Many sellers have a positive cumulative abnormal return of about 20 to
40 percent after the sale of a company (Martynova & Renneboog, 2008; Glaum & Hutzschenreuter, 2010). At the same time, around 50 percent have little negative cumulative abnormal return after the acquisition of a company. The remaining proportion has no or little positive cumulative abnormal return (Goergen & Renneboog, 2004). One possible explanation could be found in the corporate control market, which is highly competitive and also deals with the control of corporations. For each business being sold, there are many prospective buyers. Shareholders of a business up for sale can reduce the full economic value, therefore buyers cannot achieve a positive abnormal return (Glaum & Hutzschenreuter, 2010). In addition to the assessments from the view of the buyer and seller, there are studies which determine the performance of a combined entity. Investigations often conclude that, in this case, there is a significant positive abnormal return. The cumulative abnormal return (CAR) is on average two percent, calculated with an event window of three days around the announcement of an M&A transaction. This suggests a small growth of shareholder value through M&As (Buckmann, 2012). While CAR measures M&A performance on a short-term basis, some studies measure long-term performance within the capital market, making CAR an ineffective method (Ostrowski, 2007). In this case, a method called buy-and-hold-abnormal-returns (BHAR) is often used instead (Barber & Lyon, 1997). Many findings related to long-term performance result in a significant negative abnormal return, calculated with an event window from three to five years after the transaction. Nevertheless, this method of measuring the long-term success of M&As is often criticized based on the difficulty of separating the effect of the M&A from other influences (Glaum & Hutzschenreuter, 2010). Similar to market measures, it is not possible to reach consistent conclusions using accounting measures. The problem is exacerbated when using accounting measures, however, due to the difficulty of comparing studies with different balance sheet ratios and key performance indicators (Tuch & O’Sullivan, 2007). Return on equity is used most frequently as a ratio in empirical studies. Buckmann (2012) concludes that 50 percent of M&A transactions lead to a significant reduction in return on equity. Bruner (2002) examined studies using accounting measures and different key performance indicators, and found that more than half of the significant findings had a negative influence on performance. Similar results are provided by Eisenbarth (2013). In conclusion, findings resulting from the use of accounting measures are mixed. Nevertheless, the effects of M&A transactions on key performance indicators seem to be negative (Tuch & O’Sullivan, 2007; Eisenbarth, 2013). Expert surveys attract less attention in literature than market or accounting measures, because M&A transactions tend to have a higher rate of success in studies using this methodology. One reason for this could be a lack of objectivity relating to this method (Hinne, 2008; Kerler, 2000). Studies focused on using divestments to measure the performance of M&A transactions tend towards more negative findings (Eisenbarth, 2013). In fact, only one study measured a positive influence of M&A transactions on performance (Hoffmann, 1989). A statement could not be made on studies that use divestments, though, because there is not a sufficient amount of studies to do so (Picken, 2003).

To sum up the existing literature, it can be stated that it is not at all certain M&A transactions will be successful and that, if anything, they tend to have a more negative influence on success than positive.

Method

Methodology and previous meta-analyses

To measure the success of M&As, this study uses a meta-analytic procedure. Meta-analyses used to be predominantly applied in the areas of medicine and psychology but are now increasingly used in business studies (Hedges and Olkin, 1985). There are many methods of conducting a meta-analysis on a set of empirical studies, but the suitability of each different method depends to a large extend on both the research aim as well as the data, which can be obtained from the studies used in the meta-analysis. Some methods are more effective in extracting the necessary information than others (Flickinger, 2009). The more advanced types of meta-analysis, though, are based on the same key premises and have a similar approach to calculating magnitudes from studies (Guzzo et al., 1987). Hunter and Schmidt’s approach (2004) has been particularly acknowledged in economic literature (Certo et al., 2006; Lepine, Podsakoff, & Lepine, 2005). In order to be able to compare and combine all the results from the studies, they must first be adapted to the same numerical scale (Field & Gillett, 2010). Often, different studies with a hypothetical relationship measure the same construct using different variables. Because of this, it is difficult to compare the findings directly. To solve this problem, meta-analysis uses an ‘effect size’ (Flickinger, 2009). With the help of this size each study’s statistical results can be converted into a parameter that contains information about its magnitude, direction or both (Lipsey & Wilson, 2005). The type of effect size used for a meta-analysis has to be the same in all the studies that are included. The choice of effect size depends on the described relationship within the considered studies and the statistical form of the findings (Mosteller & Colditz, 1996). Commonly used effect size statistics are Cohen’s d or its modification. Hedges’ g. These effect size statistics calculate the difference between the mean of a trial and a control group (Cohen, 1977; Hedges & Olkin, 1985).

In medical and sociological research, the use of the Odds Ratio as an effect size is very common. The Odds Ratio compares the relative probability of occurrence of a condition or an occasion in two different groups (Berlin et al., 1989; DerSimonian & Laird, 1986; Fleiss, 1994). In the field of business studies, the correlation coefficient according to Pearson is the most frequently used effect size statistic (Certo et. al. 2006, 820). One reason for this could be the differences in the application method. It is uncommon to run a comparison between a trial and a control group, which is completely different from medical, psychological or sociological studies (Flickinger, 2009). In addition to effect size statistics, a meta-analysis uses another statistic to describe the research results for each individual study included. A single result of a study included in a meta-analysis is interpreted as a statistical representation of a single empirical relationship between the considered variables. The number of samples used in empirical studies that are also included in a meta-analysis is not the same, meaning that bigger samples have a smaller sampling error and therefore are of higher statistical relevancy. Meta-analyses adjust this by weighting each effect size obtained by an empirical study through its sample size. Studies which do not contain this information cannot be considered for inclusion in the meta-analysis (Flickinger, 2009).

Up until now there have been four meta-analyses dealing with the direct or indirect performance of M&As:

1. The first meta-analysis concerning the results of M&As was carried out by Datta, Pinches and Narayanan (1992).
This meta-analysis examined the influence of different factors on the shareholder value as a consequence of M&As. In this case a meta-analytical approach called replication analysis was adopted, which shows similarities to the multiple regression analysis (Flickinger, 2009; Farley, Lehmann, & Ryan, 1981). The authors drew the conclusion that selling a company is significantly profitable for the shareholders. On the contrary, the acquisition of a company leads to a significant loss for the shareholders (Datta et al., 1992).

2. King et al. (2004) employed the same theoretical construct as Datta et al. (1992). They evaluated the empirical influence of different variables on the performance of a company after an M&A transaction. For their meta-analysis, they used Hunter and Schmidt’s approach. The correlation coefficient according to Pearson is used as an effect size. The correlation here describes the link between the chosen variables and financial performance. King et al. find robust results that on average most of the applied variables do not have a positive effect on the financial performance of an M&A transaction. In fact, they have a slightly negative effect on the financial performance of an M&A transaction. As a conclusion, King et al. suspect the existence of other still unknown variables, which significantly contribute to the performance of M&A. The authors ask for further development of the theory and subject of M&A and also suggest a modification of research methods (King et al., 2004).

3. The third meta-analysis was done by Stahl and Voigt (2008). Like King et al. (2004) it uses the same approach as Hunter and Schmidt (2004). It aims specifically to research the influence of cultural difference on M&A performance. As a final conclusion Stahl and Voigt find neither a positive nor a negative effect of cultural difference on the performance of M&A. For future research the authors suggest not examining cultural differences in relation to M&A success, rather in relation to how these cultural differences affect the process of integration within an M&A transaction (Stahl & Voigt, 2008).

4. In their meta-analysis Homberg, Rost and Osterloh (2009) examine whether related transactions have a positive influence on the performance of a company that has undertaken an M&A. A related transaction occurs if M&A show similarities in relation to their business field, culture, technologies or company size. To conduct the meta-analysis the authors also use Hunter and Schmidt’s approach (2004). They conclude that related transactions do not have any positive effect on the success of an M&A (Homberg et al., 2009).

For the purpose of these analyses, the performance of M&A is only examined from the point of view of the purchasing company. The reason for this is that most empirical studies use this perspective and conducting a study from the perspective of the seller is unusual (Wirtz, 2014). Furthermore, the focus of this study is on M&A performance as a whole. So far the outlined meta-analyses have used an effect size which explains the performance of M&A on the basis of two variables. In contrast, this study applies an effect size which explains the performance of M&A on the basis of one variable. Until now it has not been very common to use an effect size with a one-variable relationship in business studies (Flickinger, 2009). According to Lipsey and Wilson (2005), two conditions must be fulfilled to successfully conduct a meta-analysis with an effect size that uses a one-variable relationship. First, all variables of the included empirical studies need to be operationalised in a similar way. Secondly, it is vital to define an effect size which contains all relevant information regarding the research question and which makes it possible to determine the standard error. In the context of this study it seems obvious to use CAR as an effect size statistic. CAR can be interpreted as share holders’ estimations of future earnings on company-specific events (Lubatkin & Shrieves, 1986). Simplified positive CARs can be understood as successful M&A transactions and negative CARs as unsuccessful M&A transactions (Glaum & Hultschenreuter, 2010). The CAR is also a widely used scale for the performance measurement of M&A. This helps to facilitate the search for relevant empirical studies for the meta-analysis (Schoenberg, 2006).

Either the arithmetic mean or the proportions can be used as effect sizes for CAR (Lipsey and Wilson, 2005). For the purpose of this meta-analysis, proportions are used as effect sizes. Arithmetic means are not suitable because it is not common to indicate the standard deviation in event studies, however the standard deviation is necessary to calculate the inverse variance weight. The conduction of the meta-analysis in this study is based on six steps according to Field and Gillett’s recommendation (2010) and uses the method created by Flickinger (2009).

The first step will be to search for empirical studies suitable for this meta-analysis (see section below on literature search). In step two, empirical studies will be selected from this search to be included in the meta-analysis (see section below on selection of literature). In step three, the effect sizes will be calculated (see section below on calculation of the effect sizes), and in step four, the meta-analysis will be conducted (see section below on the meta-analysis). Moderator analysis and publication bias are additional analyses which will be examined in step five (see section below on additional analyses). Finally, the results will be summarized, limitations portrayed and further implications given (under the summary section).

### Framework and conduction of the meta-analysis

#### Literature search

The first step is to search for primary studies with the same research question as the meta-analysis (Field & Gillett, 2010), an essential part of the process. The goal is to identify all relevant studies containing information regarding the research question. (Cooper, 2010). To guarantee a thorough and comprehensive literature search, this step must be done in a systematic way (Lipsey & Wilson, 2005; Hunter & Schmidt, 2004). In this study, the literature search is done in four phases similar to Flickinger (2009). All relevant data will be collected from the studies then evaluated and selected in the next step.

<table>
<thead>
<tr>
<th>Table 1: Overview of English and German keyword searches</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keywords</strong></td>
</tr>
<tr>
<td>M&amp;A</td>
</tr>
<tr>
<td>Akquisition/acquisition</td>
</tr>
<tr>
<td>Fusion/merger</td>
</tr>
<tr>
<td>Unternehmensübernahme/takeover</td>
</tr>
<tr>
<td>Unternehmensvereinigung</td>
</tr>
<tr>
<td>Unternehmenstransaktion</td>
</tr>
<tr>
<td>Unternehmenskauf</td>
</tr>
<tr>
<td>Source: Authors</td>
</tr>
</tbody>
</table>

www.journals.cz
In the first phase of data collection, various computerized databases are searched. Included are the following computerized databases: WISO, EconBiz, Business Source Premier and JSTOR. All of these databases operate in the same way, using an extensive list with keywords created to cover the entire research area. The search for this study was done with the help of keywords such as “M&A” or “acquisition” and suitable additions e.g. “study” or “performance”. German as well as English studies were considered in this search. An overview of English and German keywords used in this study can be found in Table 1. After the database search, a manual journal search was conducted. The reason behind this was to find further matches in relation to the research area (Arendt, 2007). In Table 2, the ten manually researched journals are listed.

In the third phase, a backwards and forwards search is applied. That entails previously found studies being examined in their reference section for further relevant empirical studies. The forwards search uses the scientific citation indexing service Web of Science to find studies which cite relevant research papers, which have already been found (White, 2009). The fourth and final phase of the literature search aims to find unpublished studies, as literature on the subject of meta-analysis often recommends including unpublished studies in meta-analyses (Field & Gillett, 2010). Empirical studies with significant results are eight times more likely to be published than studies with non-significant results (Greenwald, 1975). This is because researchers rarely submit studies with non-significant results, and also because publishers tend to reject such studies (Hedges, 1984; Dickersin, Min, & Meinert, 1992). It is assumed that meta-analyses which do not include unpublished study results have a publication bias, also known as a file-drawer problem. This is a systematic cause by selective publishing practices. To avoid this kind of publication bias, the database of the Social Science Research Network (SSRN) was searched for unpublished studies related to the performance of M&A (Flickinger, 2009).

Selection of literature

Previous investigations showed that the major source of differences between meta-analyses with the same research question was their selection of literature (Wanous, Sullivan, & Mitchell, 1992). This statement reflects the importance of the second step of meta-analyses. The studies selected for this meta-analysis using the literature search (see section on literature search) had to fulfill the following four criteria (Flickinger, 2009):

1. Only studies published from the year 2004 onwards are used within this meta-analysis. This should guarantee a current overview of the performance of M&A.
2. All empirical studies used in this meta-analysis have to measure the reaction of the capital market in CAR as a consequence of an M&A announcement to be included.

Additionally, the number of all positive CARs in percent or other statistics from which this could be calculated have to be mentioned.

3. The event window by which the CAR is measured should be close to the day of the M&A announcement. previous empirical studies showed that a small event window usually gives the best representation of the significant moment of an event (Dann, Mayers, & Raab Jr., 1977; Mitchell & Netter, 1989; Ryngaert & Netter, 1990). In addition, small event windows minimize the risk of including other events that do not reflect the reaction of the capital market to M&A transactions (Lubatkin & Shrieves, 1986). Most samples used in this meta-analysis have an event window of three days from day -1 to day +1 after the announcement of the transaction. In principle, only studies with an event window no longer than 41 days (from day -20 to day +20 after the announcement of the transaction) were used.

4. In order to guarantee statistical independence only one effect size per sample size should be used during the meta-analysis (Lipsy & Wilson, 2005). This means that only those sample sizes with the highest statistical significance are used in the meta-analysis if studies have CARs for more than one event window. In contrast, some studies use different sample sizes e.g. CARs for domestic and for cross-border M&As. Here, one effect size is used per sample size within the meta-analysis (Flickinger, 2009).

<table>
<thead>
<tr>
<th>Authors</th>
<th>Journal</th>
<th>Period</th>
</tr>
</thead>
</table>

Source: Authors

Table 2: Manually searched journals

<table>
<thead>
<tr>
<th>Journal</th>
<th>Source: Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Financial Management</td>
<td></td>
</tr>
<tr>
<td>IIE Transactions</td>
<td></td>
</tr>
<tr>
<td>International Review of Financial Analysis</td>
<td></td>
</tr>
<tr>
<td>Journal of Business Economics</td>
<td></td>
</tr>
<tr>
<td>Journal of International Business Studies</td>
<td></td>
</tr>
<tr>
<td>The Journal of Finance</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Overview of studies included
In this meta-analysis 35 effect sizes have been calculated using 33 empirical studies. A total of 55,399 M&A transactions are represented by this meta-analysis, all taking place between 1950 and 2010. Table 3 shows all studies included.

Coding scheme

All necessary information was extracted from each selected study, including statistical results, sample sizes and characteristics which could have an influence on the effect size (Guzzo et al., 1987). The coding scheme of meta-analyses consists of two parts: encoding information regarding the empirical results of the studies (effect size) and encoding additional information about the study’s characteristics (study descriptor) (Lipsey and Wilson, 2005). In order to calculate the effect size of each sample, the percentage of positive CARs had to be taken from the studies. Additionally, the sample size was noted for each sample. The sample size is a further requirement for calculating the inverse variance weights of each effect size (Flickinger, 2009). Moreover, the following additional information about the individual studies was collected in order for it to be used in the moderator analysis:

- Type of M&A transaction. The sample sizes were distinguished by domestic and cross-border M&A.
- Year of publication of the study. In order to identify possible variations regarding time of publication, the year of publication of each study had to be codified.
- Event window of the CAR. The different event windows used for calculating the CAR were taken into account. The 35 sample sizes extracted from the 33 studies were measured during 11 different event windows. 14 sample sizes reported CAR from day -1 to day +1 around the announcement date, while 4 sample sizes reported CAR on the announcement date (0) and between day -2 to day +2. The remaining sample sizes spread across event windows between day -20 to day +20 (see Table 4).

Calculation of the effect sizes

As previously mentioned, the effect sizes in this study are calculated using proportions, which can have any value between 0 and 1 (Lipsey and Wilson, 2005). In empirical studies, they are usually expressed directly or as a percentage (Pauser, 2007; Bassen, Schiereck, & Wübben, 2010; Danbolt & Maciver, 2012). The percentage is divided by 100 to be converted into a proportion. If there is no proportion or percentage mentioned in a study, the proportion can be determined by the quantity of transactions with a positive CAR divided by the quantity of the total sample size (Card, 2012; Rani Yadav, & Jain, 2014; Kirchhoff, Schiereck, & Mentz, 2006). According to Lipsey and Wilson (2005), the proportion is suitable for estimating the mean proportion of all the selected studies. Nevertheless, an eye should be kept on the fact that the size of the confidence interval around the mean effect size is underestimated, and that the degree of heterogeneity exceeding effect sizes is overestimated (especially if the value of proportions is less than 0.2 or more than 0.8). To avoid this problem, it is useful to undertake a 'logit transformation' (Lipsey and Wilson, 2005; Koch and Windeler, 1999). During a logit transformation the calculated proportions of the empirical studies are transformed into a logit value (Card, 2012).

Table 5: Overview of the effect size per sample

<table>
<thead>
<tr>
<th>Sample of</th>
<th>p</th>
<th>n</th>
<th>ESi</th>
<th>SEi</th>
<th>wi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akbult and Matsusaka (2010)</td>
<td>0.39</td>
<td>3,473</td>
<td>-0.447</td>
<td>0.035</td>
<td>826.227</td>
</tr>
<tr>
<td>Asimakopoulos and Athanasoglou (2013)</td>
<td>0.503</td>
<td>145</td>
<td>0.014</td>
<td>0.166</td>
<td>36.248</td>
</tr>
<tr>
<td>Baker et al. (2012)</td>
<td>0.549</td>
<td>1,066</td>
<td>0.197</td>
<td>0.062</td>
<td>263.941</td>
</tr>
<tr>
<td>Bassen et al. (2010)</td>
<td>0.62</td>
<td>78</td>
<td>0.49</td>
<td>0.233</td>
<td>18.377</td>
</tr>
<tr>
<td>Beitel et al. (2004)</td>
<td>0.49</td>
<td>98</td>
<td>-0.041</td>
<td>0.202</td>
<td>24.49</td>
</tr>
<tr>
<td>Bouzaglou and Navatte (2012)</td>
<td>0.51</td>
<td>120</td>
<td>0.04</td>
<td>0.183</td>
<td>29.988</td>
</tr>
<tr>
<td>Danbolt and Maciver (2012)</td>
<td>0.466</td>
<td>146</td>
<td>-0.136</td>
<td>0.166</td>
<td>36.331</td>
</tr>
<tr>
<td>Danbolt and Maciver (2012)</td>
<td>0.323</td>
<td>146</td>
<td>-0.74</td>
<td>0.177</td>
<td>31.926</td>
</tr>
<tr>
<td>Ebneleth and Theuven (2007)</td>
<td>0.548</td>
<td>31</td>
<td>0.194</td>
<td>0.361</td>
<td>7.677</td>
</tr>
<tr>
<td>Faley (2011)</td>
<td>0.49</td>
<td>2,457</td>
<td>-0.04</td>
<td>0.04</td>
<td>614.004</td>
</tr>
<tr>
<td>Gaur et al. (2013)</td>
<td>0.548</td>
<td>1,074</td>
<td>0.194</td>
<td>0.061</td>
<td>265.984</td>
</tr>
<tr>
<td>Gleason et al. (2005)</td>
<td>0.598</td>
<td>97</td>
<td>0.397</td>
<td>0.207</td>
<td>23.32</td>
</tr>
<tr>
<td>Gubbi et al. (2010)</td>
<td>0.586</td>
<td>418</td>
<td>0.348</td>
<td>0.099</td>
<td>101.4</td>
</tr>
<tr>
<td>Guest (2009)</td>
<td>0.551</td>
<td>851</td>
<td>0.205</td>
<td>0.069</td>
<td>210.526</td>
</tr>
<tr>
<td>Jansen et al. (2012)</td>
<td>0.51</td>
<td>18,872</td>
<td>0.038</td>
<td>0.015</td>
<td>4,716.26</td>
</tr>
<tr>
<td>Kengelbach et al. (2012)</td>
<td>0.451</td>
<td>20,975</td>
<td>-0.197</td>
<td>0.014</td>
<td>5,193.39</td>
</tr>
<tr>
<td>Kirchhoff and Schiereck (2011)</td>
<td>0.413</td>
<td>104</td>
<td>-0.35</td>
<td>0.199</td>
<td>25.221</td>
</tr>
<tr>
<td>Kirchoff et al. (2006)</td>
<td>0.44</td>
<td>50</td>
<td>-0.241</td>
<td>0.285</td>
<td>12.32</td>
</tr>
<tr>
<td>Kymaz and Baker (2008)</td>
<td>0.414</td>
<td>869</td>
<td>-0.346</td>
<td>0.069</td>
<td>210.868</td>
</tr>
<tr>
<td>Laabs (2009)</td>
<td>0.59</td>
<td>206</td>
<td>0.364</td>
<td>0.142</td>
<td>49.831</td>
</tr>
<tr>
<td>Lee and Chung (2013)</td>
<td>0.398</td>
<td>1,036</td>
<td>-0.413</td>
<td>0.063</td>
<td>248.285</td>
</tr>
<tr>
<td>Lehn and Zhao (2006)</td>
<td>0.402</td>
<td>714</td>
<td>-0.397</td>
<td>0.076</td>
<td>171.643</td>
</tr>
<tr>
<td>Lensink and Maslennikova (2008)</td>
<td>0.547</td>
<td>75</td>
<td>0.187</td>
<td>0.232</td>
<td>18.587</td>
</tr>
<tr>
<td>Masulis and Mobb (2011)</td>
<td>0.56</td>
<td>205</td>
<td>0.242</td>
<td>0.141</td>
<td>50.505</td>
</tr>
<tr>
<td>Papadakis and Thanos (2010)</td>
<td>0.48</td>
<td>50</td>
<td>-0.08</td>
<td>0.283</td>
<td>12.48</td>
</tr>
<tr>
<td>Pauser (2007)</td>
<td>0.585</td>
<td>106</td>
<td>0.343</td>
<td>0.197</td>
<td>25.736</td>
</tr>
<tr>
<td>Piskula (2011)</td>
<td>0.366</td>
<td>216</td>
<td>-0.551</td>
<td>0.141</td>
<td>50.106</td>
</tr>
<tr>
<td>Rani et al. (2014)</td>
<td>0.68</td>
<td>228</td>
<td>0.753</td>
<td>0.142</td>
<td>49.627</td>
</tr>
<tr>
<td>Rani et al. (2014)</td>
<td>0.564</td>
<td>225</td>
<td>0.259</td>
<td>0.134</td>
<td>55.316</td>
</tr>
<tr>
<td>Rani et al. (2013)</td>
<td>0.6</td>
<td>155</td>
<td>0.405</td>
<td>0.164</td>
<td>37.2</td>
</tr>
<tr>
<td>Schoenberg (2006)</td>
<td>0.5</td>
<td>61</td>
<td>0</td>
<td>0.256</td>
<td>15.25</td>
</tr>
<tr>
<td>Sears and Hoetker (2014)</td>
<td>0.361</td>
<td>97</td>
<td>-0.572</td>
<td>0.211</td>
<td>22.371</td>
</tr>
<tr>
<td>Shahhr and Venkateswaran (2009)</td>
<td>0.393</td>
<td>816</td>
<td>-0.433</td>
<td>0.072</td>
<td>194.727</td>
</tr>
<tr>
<td>Thomas (2008)</td>
<td>0.574</td>
<td>61</td>
<td>0.297</td>
<td>0.259</td>
<td>14.918</td>
</tr>
<tr>
<td>Wübben (2007)</td>
<td>0.51</td>
<td>78</td>
<td>0.04</td>
<td>0.227</td>
<td>19.492</td>
</tr>
</tbody>
</table>

Source: Authors

The logit effect size of proportions ESi is calculated by

\[
ES_i = \log_p \left[ \frac{p}{1-p} \right]
\]  

with p as the proportion of the positive CAR to the particular sample size. i stands for the particular sample from 1 to k.

Table 4: Samples per event window

<table>
<thead>
<tr>
<th>Event Window</th>
<th>[0;+1]</th>
<th>[0;+5]</th>
<th>[0]</th>
<th>[-1;+1]</th>
<th>[-1;+5]</th>
<th>[-2;+2]</th>
<th>[-3;+3]</th>
<th>[-5;+5]</th>
<th>[-10;+10]</th>
<th>[-20;+20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Source: Authors

www.journals.cz
The mean effect size \( ES \) is supplemented by the following and Wilson, 2005): effect size \( ES \) can be converted back to proportions (Lipsey and Wilson, 2005). To facilitate the interpretation of the final results the mean effect size \( \bar{ES} \) is weighted by its inverse variance weight \( w_i \) of each effect size is given by:

\[
w_i = \frac{1}{SE_i^2} = n \cdot p \cdot (1 - p)
\]  

(3)

Table 5 shows the proportion, sample size, effect size \( ES_i \), standard error \( SE_i \) and the inverse variance weight \( w_i \) of each sample included in this meta-analysis.

The meta-analysis

In the following section the results of each sample will be compiled for the meta-analysis. Each effect size \( ES_i \) is weighted by its inverse variance weight \( w_i \) to calculate the weighted mean effect size \( \bar{ES} \) (Lipsey and Wilson, 2005). According to this, the formula for the weighted mean effect size \( \bar{ES} \) is:

\[
\bar{ES} = \frac{\sum (w_i \cdot ES_i)}{\sum w_i}
\]  

(4)

To facilitate the interpretation of the final results the mean effect size \( \bar{ES} \) can be converted back to proportions (Lipsey and Wilson, 2005):

\[
p = \frac{e^{logit}}{e^{logit} + 1}
\]  

(5)

The mean effect size \( \bar{ES} \) is supplemented by the following statistics:

- **Confidence intervals around the mean effect size \( \bar{ES} \).** The range in which the population mean is most likely to be situated is indicated by the confidence intervals (Lipsey and Wilson, 2005). Normally it is not possible for researchers to collect all the data for a population. Therefore, they focus on samples which include population subsets and try to closely display the population (Rasch et al., 2010). For example, a 95 percent confidence interval shows that the population mean is up to 95 percent in between those two values. For the mean effect size \( \bar{ES} \) the confidence interval is based on the standard error and a critical value of the \( z \)-distribution. To calculate the standard error for the mean effect size \( SE_{\bar{ES}} \), the root of 1 divided by the sum of all inverse variance weights \( w_i \) must be found:

\[
SE_{\bar{ES}} = \frac{1}{\sum w_i}
\]  

(6)

For \( z = 0.001 \) that represents the desired level of confidence. If the product of this is subtracted from the mean effect size \( \bar{ES} \), the lower limit \( ES_L \) is obtained. If the product is added to the mean effect size \( \bar{ES} \), the upper limit \( ES_U \) can be acquired.

\[
ES_L = \bar{ES} - \frac{z_{(1-\alpha)}}{\bar{SE}_{\bar{ES}}}
\]  

(7)

\[
ES_U = \bar{ES} + \frac{z_{(1-\alpha)}}{\bar{SE}_{\bar{ES}}}
\]  

(8)

If 0 is not included in the confidence interval, the mean effect size \( \bar{ES} \) is statistically significant at \( p \leq \alpha \).

A direct test of the statistical significance of the mean effect size \( \bar{ES} \) can be taken with a \( z \)-test. The \( z \)-test is given by:

\[
z = \frac{\bar{ES}}{SE_{\bar{ES}}}
\]  

(9)

\( |\bar{ES}| \) is the absolute value of the mean effect size \( \bar{ES} \). The result of the \( z \)-test is distributed as standard normal variate. The result can be interpreted as statistically significant with \( p \leq 0.05 \) two-tailed if \( z \) is greater than or equal to 1.96. If \( z \) is greater than or equal to 2.58, it is statistically significant with \( p \leq 0.01 \) two-tailed. When \( z \) is greater than or equal to 3.29, it is statistically significant with \( p \leq 0.001 \) two-tailed (Lipsey and Wilson, 2005).

- **Significance test of heterogeneity**

During meta-analysis it is important that the different effect sizes \( ES_i \) estimate the same population effect size. In a homogenous distribution the only difference between the several effect sizes \( ES \) and the population effect size exists in the sampling error. If a statistical test rejects the null hypothesis of homogeneity it means that the variability of effect sizes is greater than was predicted by the sampling error (Lipsey and Wilson, 2005). The heterogeneity of effect sizes is often evaluated by a \( Q \) statistic. This test is called the test of homogeneity; there is also a lesser-known test of heterogeneity (Card, 2012). It is distributed as chi-square with \( k-1 \) degrees of freedom with \( k \) representing the number of effect sizes (or samples). \( df(Q) \) is the degree of freedom. The formula for \( Q \) is given by (Lipsey and Wilson, 2005):

\[
Q = \sum w_i \cdot (ES_i - \bar{ES})^2
\]  

(10)

Table 6 summarizes the results of the meta-analysis.

<table>
<thead>
<tr>
<th>Number of samples ( k )</th>
<th>Lower limit ( \bar{ES}_L )</th>
<th>Upper limit ( \bar{ES}_U )</th>
<th>( z )-value</th>
<th>( Q )-value</th>
<th>( df(Q) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>-0.07</td>
<td>-0.114</td>
<td>-0.08</td>
<td>11.343 ***</td>
<td>523.448 ***</td>
</tr>
</tbody>
</table>

* \( p < 0.05 \), ** \( p < 0.01 \), *** \( p < 0.001 \)

Source: Authors

The logit point estimate, which is equal to the mean effect size \( \bar{ES} \), has a value of -0.097. Converted back into proportions, 47.6 percent of all M&A announcements receive a positive reaction on the capital market. The result is highly
significant (p < 0.001). The Q value is also highly significant (p < 0.001) and can be interpreted as a heterogeneity distribution. If effect sizes are heterogenic, further investigations are recommended to find the source of heterogeneity (Lipsey and Wilson, 2005).

Additional analyses

Moderator analysis

If there is a remarkable heterogeneity in the effect sizes among the meta-analysis samples, it is useful to search for the reasons behind this heterogeneity. For meta-analyses, this can be carried out with the help of a moderator analysis which tries to explain the heterogeneity of the effect sizes between the samples using different coded study characteristics as predictors. The aim of a moderator analysis is to identify study characteristics which have a positive or negative impact on the effect size (Card, 2012). To measure the effects, the moderators have categorized empirical studies according to their methodological or substantive attributes and for each category, a subgroup meta-analysis must be conducted (Cortina, 2003). This type of meta-analysis is used if the moderator variables are categorical rather than continuous (Card, 2012; Flickinger, 2009). All calculations are based on the fixed-effects model, in which it is assumed that the effect sizes of the empirical studies estimate the effect on the population and only differ by the sampling error. The generalization of the results towards a population is only possible if the sample is homogenous. If this cannot be guaranteed, it should be supplemented by a random-effects model. The random-effects model assumes that single empirical studies are taken out of the population randomly and that the effect sizes are distributed, in general normally. The difference between these two models is that the random-effects model assumes, in addition to the sampling error, that there is a random error representing the deviation between empirical studies. Along with the inverse variance weight w, and the mean effect size $\bar{ES}$, the confidence interval changes with the random-effects model (Lipsey and Wilson, 2005). The fixed-effects model is the model accepted in practice, primarily because it can be put into operation more easily but also because it is founded on a wider theoretical base than the random-effects model (Hunter and Schmidt, 2004).

To test moderators according to their categories in meta-analyses, groups of studies are compared then organised according to the status of selected categorical moderators (Card, 2012). The formula for Q and for the degrees of freedom df (Q) will be recalculated with the addition of the term “total”. This is meant to achieve a clearer classification and should also constitute the total heterogeneity between the effect sizes.

$$Q_{\text{total}} = \sum w_i \cdot (ES_i - \bar{ES})^2 \quad (11)$$

$$df (Q)_{\text{total}} = k - 1 \quad (12)$$

To test the categorical moderators, the total heterogeneity ($Q_{\text{total}}$) is separated into two components. The component $Q_{bw}$ tests the heterogeneity between the groups and the component $Q_{in}$ tests the heterogeneity within the groups (Bortz and Döring, 2006). The formula is given by

$$Q_{\text{total}} = Q_{bw} + Q_{in} \quad (13)$$

The essential question when evaluating categorical moderators is whether a larger heterogeneity between the groups is to be expected. If this condition is met, it can be concluded that the groups of studies organised according to categorical moderators, are different from each other and, consequently, the categorical moderator is reliable as far as the effect sizes are concerned.

If this condition is not met, no significance for this moderator exists and the moderator does not have any influence on the effect sizes of the studies.

The easiest way to determine the heterogeneity between the groups ($Q_{bw}$) is to rearrange the formula: ($Q_{bw} = Q_{\text{total}} - Q_{in}$). Only the heterogeneity within the groups ($Q_{in}$) needs to be calculated to obtain the heterogeneity between the groups ($Q_{bw}$). To calculate the heterogeneity within a group $g$, the following formula applies:

$$Q_g = \sum w_i \cdot (ES_i - \bar{ES}_g)^2 \quad (14)$$

$ES_g$ is the total effect size between the samples within group g. The degree of freedom of group g, df ($Q_g$), is calculated by

$$df (Q_g) = k_g - 1 \quad (15)$$

$k_g$ is the number of samples within group g. To achieve heterogeneity within the groups ($Q_{in}$) it is summed up for each group g ($Q_g$) using the following:

$$Q_{in} = \sum_{g=1}^{G} Q_g \quad (16)$$

$$df(Q_{in}) = \sum_{g=1}^{G} df(Q_g) = k - G \quad (17)$$

G represents the number of groups and $df (Q_{in})$ represents the degrees of freedom within the group (Card, 2012).

The following moderator variables are used in this study:

- Type of M&A (domestic or cross-border M&As)
- Year of publication of the study (before 2010/after 2010)
- Event window for the CAR

These variables serve as a means of examining the influence of additional study characteristics on the meta-analysis. Each single variable was tested in a separate moderator analysis.

1. Type of M&A

The first moderator analysis examines the type of M&A. Table 7 summarizes the results. The type of M&A is separated into two groups: domestic M&As and cross-border M&As. Because this information is not given in all the studies included in the meta-analysis, 12 out of 35 samples had to be excluded from this moderator analysis (Asimakopoulos & Athanasoglu, 2013; Bouzarrou & Navatte, 2012; Gleason, McNulty, & Pennathur, 2005; Kengelbach, Roos, & Klemmer, 2013; Kirchhoff & Schiereck, 2011; Kirchhoff et al., 2006; Laabs, 2009; Lehn & Zhao, 2006; Lensink & Maslennikova, 2008; Masulis & Mobbs, 2011; Pauser, 2007; Sears & Hoetker, 2014).
The logit point estimate of -0.045 for the group of domestic M&As is clearly different from the logit point estimate of 0.287 for cross-border M&As. Converted back to proportions, only 48.9 percent of domestic M&As and 57.1 percent of cross-border M&As receive a positive reaction from the capital market after the announcement. There is no overlap between the 95 percent confidence interval for the two moderator variables, as point estimates are highly significant (p < 0.001). The Q-value between the groups Q_{bw} is 30.498 is also highly significant (p < 0.001). This indicates a heterogeneous distribution of effect sizes between the two groups. The Q-value within the group Q_{gw} is heterogeneous at p < 0.001 for domestic M&As and heterogeneous at p < 0.01 for cross-border M&As. Despite the confirmed moderators (domestic M&As and cross-border M&As) the Q-statistic within the groups remains heterogeneous. Reasons for this could be different branches or country-specific types of industrialization within the groups (Kengelbach et al., 2012). This could be a starting point for further research.

2. Year of publication of the studies (before and after 2010)

Table 8 provides the results for the moderators according to the year of publication of studies included in the meta-analysis. 15 samples were extracted from studies published before 2010 and 20 samples were extracted from studies published during or after 2010.

The point estimates of the two moderator variables are slightly different. The 95 percent confidence interval is overlapped. After being converted into proportions, studies published before 2010 received a positive reaction of 46.5 percent from the capital market after the announcement of an M&A transaction, while studies published after 2010 received a positive reaction of 47.7 percent. The Q-value between the groups Q_{bw}, however, is not statistically significant (p < 0.05). There is no significant heterogeneity between studies published before and after the year 2010.

3. Event window for the CAR

The different event windows for the calculation of the CAR from the studies included in the meta-analysis are the subject of the last moderator analysis. Five groups were formed, which were then separated into groups of one-day window, two-day window, three-day window, five-day window and an event window of six or more days. Table 9 provides the results for this moderator analysis.

The point estimates for the groups are between -0.169 and -0.007, the former value corresponding to the group with an even window lasting longer than five days and the latter to the group with a three-day window. Converted back to proportions, the groups show a positive reaction between 45.8 and 49.8 percent to the announcement of M&A transactions on the capital market. The Q-value between the groups Q_{gw} is 140.129 and is highly significant (p < 0.001). This indicates a heterogeneous distribution of effect sizes between the five groups. According to the Q-value within the groups Q_{gw} only the group with a five-day window is homogenous. One reason for the heterogeneity of the other groups could be differences in the quality of the individual empirical studies (Bortz and Döring, 2006).

Publication bias

Publication bias can also be analysed. As mentioned before, journals as well as researchers tend to publish empirical studies with significant results rather than those without, leading to a publication bias. Rosenthal (1979) tries to overcome this issue with the so-called Fail-Safe-N. The Fail-Safe-N tries to answer the question of how many non-significant studies are needed to statistically avoid the significant overall effect (Orwin, 1983; Bortz and Döring, 2006).

The Fail-Safe-N is given by

$$N_k = \frac{\left( \sum_{i=1}^{k} z_i^2 \right)}{2.706} - k$$

(18)

$$\sum_{i=1}^{k} z_i$$ is sum of the single z-values of all studies k included in the meta-analysis (Field and Gillett, 2010).

For this meta-analysis, Rosenthal’s Fail-Safe-N (N_{k}) is 3,938, meaning 3,938 unpublished studies with a non-significant result are necessary to avoid a significant overall effect.

CONCLUSION

This paper measured the success of M&A transactions using a meta-analysis. Meta-analyses are instruments that aggregate results of several empirical studies by employing a transparent and systematic approach and using comprehensible statistical methods. Furthermore, meta-
analyses represent key study findings in a more sophisticated and differentiated way than narrative methods. A meta-analysis is a variable sensitive method because the effect size is calculated for each included sample. In addition, it is possible to use further factors (moderator analyses) that cannot be examined by primary studies. Last but not least the meta-analysis is an organized approach that can handle almost unlimited information about research results from empirical studies (Lipsey and Wilson, 2005).

The results of this paper suggest that less than half (47.6 percent) of all M&A transactions worldwide are successful. The performance of a company after an M&A is measured using CAR, which expresses the reaction of the capital market to the announcement of M&A transactions. The hypothesis that more than 50 percent of all M&A transactions are not successful can be confirmed. Further analyses (moderator analyses) found that cross-border transactions (57.1 percent) result in more positive reactions from the capital market than domestic M&A transactions (48.9 percent). The results differ in relation to the event window used to calculate the CAR. No heterogeneity was found concerning the year of publication of the empirical studies used, neither before 2010 nor after. This study has a number of limitations; the use of the meta-analysis method as well as the use of event studies to measure M&A performance has several disadvantages. First of all, the inclusion of event studies in the meta-analysis means their weaknesses are brought with them (Flickinger, 2009). McWilliam and Siegel (1997) examined some of these weaknesses and concluded that event studies cannot isolate the effect caused by the announcement of an M&A transaction from other effects. To overcome this problem this study only included event studies with relatively short event windows (Flickinger, 2009). Moreover, it is possible for the capital market to predict an event (here an M&A transaction) before it is publicly announced due to information leaks (McWilliams and Siegel, 1997). Lubatkin and Shriever (1986) criticize the methodology of event studies for being in conflict with the field of strategic management. Within a company the strategic management is responsible for the needs of all different stakeholders, but the performance of a company is evaluated differently depending on the group of stakeholders. Event studies imply that the only group of importance is the group of shareholders. Secondly, the expertise and additional time needed to conduct a meta-analysis is, in comparison to other conventional qualitative research reviews, a further limitation (Lipsey and Wilson, 2005). A methodological problem of meta-analyses gives rise to a further limitation. As is common, it was not possible to include all studies in this meta-analysis that measure the success of M&A transactions using the capital market’s reaction after the announcement of the agreement. The reason for this is the fact that the necessary information was not available in all the studies considered in this meta-analysis. An additional problem is publication bias, which can result in overvalued findings (Flickinger, 2009). The results of a meta-analysis are not only dependent on its transparent and systematic approach but also on the studies included: “even one red sock (bad study) amongst the white clothes (good studies) can ruin the laundry” (Field and Gillett, 2010). In conclusion, a meta-analysis is only as good as the empirical studies it uses. According to Lipsey and Wilson (2005), the greatest disadvantage is summarizing different studies into one overall effect. There are many different ways of obtaining findings for a specific research area. One example is the different methods applied to measure the M&A success. If all these findings we recombined it would be like comparing apples and oranges, so in order to avoid this limitation this meta-analysis only took proportions of positive CARs from event studies. The evaluation of the results of this meta-analysis have several implications. First of all, the analysis proves the necessity of sound risk management, which controls the success of factors influencing an M&A deal. Such a risk-conscious approach may convince the capital market that it is less likely a certain company will make the mistakes that cause low success rates of M&As during its own transactions. In relation to the low success rate, it is recommended that a careful evaluation of the target be carried out and that there be a high level of discipline during the negotiations. It is highly probable that accepting high multiples and very optimistic assumptions in the evaluation will lead to a destruction of shareholder value. In such cases the risk management has to prove its effectiveness and press for an improvement of the conditions or for an ultimate withdrawal from the deal.

References


