nature REVIEWS DRUG DISCOVERY

SUPPLEMENTARY INFORMATION
In format as provided by the authors

Demystifying industry–academia collaboration

Thomas Gersdorf, Vivianna Fang He, Ann Schlesinger, Guido Koch, Dominic Ehrismann, Hans Widmer and Georg von Krogh

https://doi.org/10.1038/d41573-019-00001-2

Supplementary Box 1 | **Data source and methodology for the qualitative analysis**

To systematically analyse the challenges in industry–academia collaborations, we employed a thematic content analysis approach to inductively analyse qualitative data from the projects in our sample. Thematic analysis helps to identify and report patterns—so-called *themes*—within qualitative data to describe phenomena[1]. The open-ended inquiry in this approach is well suited to identify similarities and differences in a data set and point to its key features[1].

Data source. We analysed the challenges reported in a survey by the project members of 187 industry–academia collaborations between the Novartis Institutes for BioMedical Research (NIBR) and academic organisations. Novartis, as the second-largest seller of prescription drugs worldwide in 2016[2], is a good example of a large company in the pharmaceutical industry engaging in industry-academia collaboration. NIBR is the research unit of Novartis and is comprised of approximately 6000 scientists, physicians and business professionals[3]. Novartis spent the second-highest amount on research and development of all companies in the industry in 2016[2], and published the largest number of scientific articles in peer-reviewed journals of all pharmaceutical companies in 2017[4].

The projects in our sample are collaborative research projects that yield a number of different outcomes including new assays, new methods, pre-validated targets, compounds, intangible learning, and knowledge transfer. These projects involved scientists from Novartis and one or more partnering academic institutions, predominantly universities and hospitals. Overall, our empirical data includes responses from 669 individuals: 416 at NIBR and 253 at the academic institutions.

Data collection. The data collection proceeded in multiple steps (see Figure S1 for an overview). First, we compiled a list of all industry–academia collaborations at NIBR that (i) could potentially include collaborative research or intellectual exchange and (ii) were active in September 2015 (the point of selection), drawn from a list of all contracts with external parties. Second, for all projects, we approached the project leader at NIBR to confirm whether the project indeed included collaborative research or intellectual exchange, and was active. Third, we asked the NIBR project leader to explicitly name all project team members from both NIBR and the partnering academic institution who were actively involved in this project. The NIBR project leader decided if the project team members from the academic partner were to be included (academics for all but 9 projects were included). As a result, we identified 783 scientists involved in 209 collaborations.

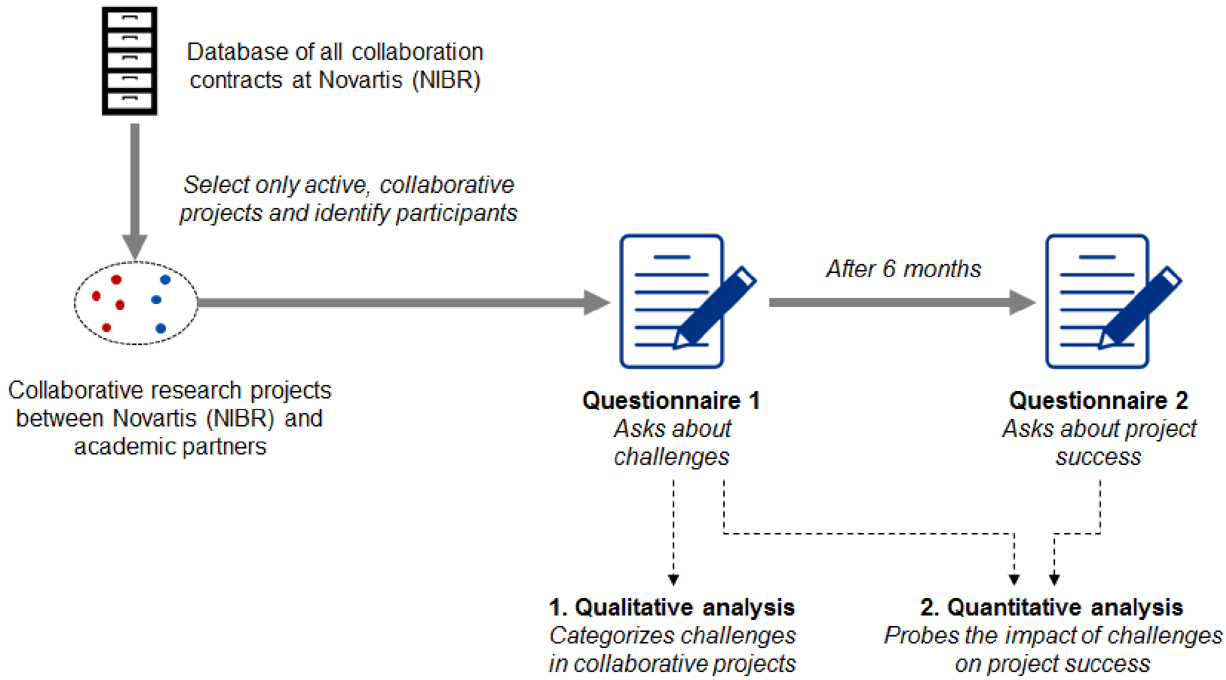

Figure S1 | **Procedure of this study.** After selecting active, collaborative projects from among all legal contracts at NIBR, we identified the members of these projects. We then sent two consecutive questionnaires to the project members from NIBR and the academic partner. The resulting data form the basis for the qualitative data analysis and the quantitative data analysis (see Supplementary Information 2 (box)).

Fourth, we sent a first questionnaire to the scientists at NIBR and the identified academic partners using an electronic survey software (Qualtrics) and an accompanying email that explained the purpose of the survey. The NIBR project leader also contacted their academic counterparts to explain the study and encourage participation. We reminded project members who did not respond with up to three emails and a telephone call. In total, 669 individuals (416 NIBR project members, 253 academic project members) in 187 projects completed the questionnaire. Fifth, about 6 months after the first questionnaire, we sent all respondents of the first questionnaire a second follow-up questionnaire that probed the progress and the outcomes of the project so far among other variables not used in this analysis. The data from this second questionnaire were used in the quantitative analysis (see Supplementary Information 2).

The projects in our sample covered all research disciplines at NIBR, and lasted on average 2.9 years. We obtained one or more responses from the academic project team members for 77% of all projects in our sample. Respondents indicated that they spent on average 24% (NIBR participants) and 26% (academic participants) of their work time on the collaborative project that the survey asked about.

Table S1 | **Response and analysis statistics** [a]

	Number of data points
Number of text fields from surveys with responses	879
- Number of blank responses	55
Number of responses considered for coding	824
- Number of ambiguous responses, not coded	58
- Number of responses indicating absence of challenges, not coded	44
Number of responses successfully coded (428 responses from Novartis participants, 294 from academic participants)	722

[a] The table shows the number of responses obtained from the first questionnaire in response to the question: "Thinking about this collaborative project as a whole, what were/are three highest hurdles in this specific project?". Details of the stratification of the data are described in the text.

Qualitative data analysis. The qualitative analysis is based only on data from the first questionnaire. In this analysis, the focus is on the responses to the open-ended question "Thinking about this collaborative project as a whole, what were/are three highest hurdles in this specific project?" Respondents were asked to name up to three challenges in three free text boxes in the questionnaire. The text boxes were not mandatory to conclude the questionnaire, and of the 669 returned questionnaires, 293 survey participants answered this question, providing 824 text-based responses (non-empty data from all three free text fields combined together, see table S1 for details). The qualitative data analysis proceeded in four steps. First, we corrected spelling mistakes, cleaned the data, combined the responses from all three text boxes responses and anonymized all 824 responses to ensure an unbiased analysis.

Second, we developed a coding scheme that categorizes the text-based responses. Coding schemes are a key element in the inductive analysis of qualitative data[5]. The development of a coding scheme involves grouping responses that concern the same topic into so-called "themes" and subsequently interpreting these themes to aggregate them into more abstract categories[5].

To develop the coding scheme, an independent research assistant who is not part of the author team was tasked with examining the data and grouped all 824 responses into themes. The development of this coding scheme was guided by the principles (i) maximize the differences between themes and (ii) minimize the difference within themes, such that the responses can be mapped unambiguously to a single theme.

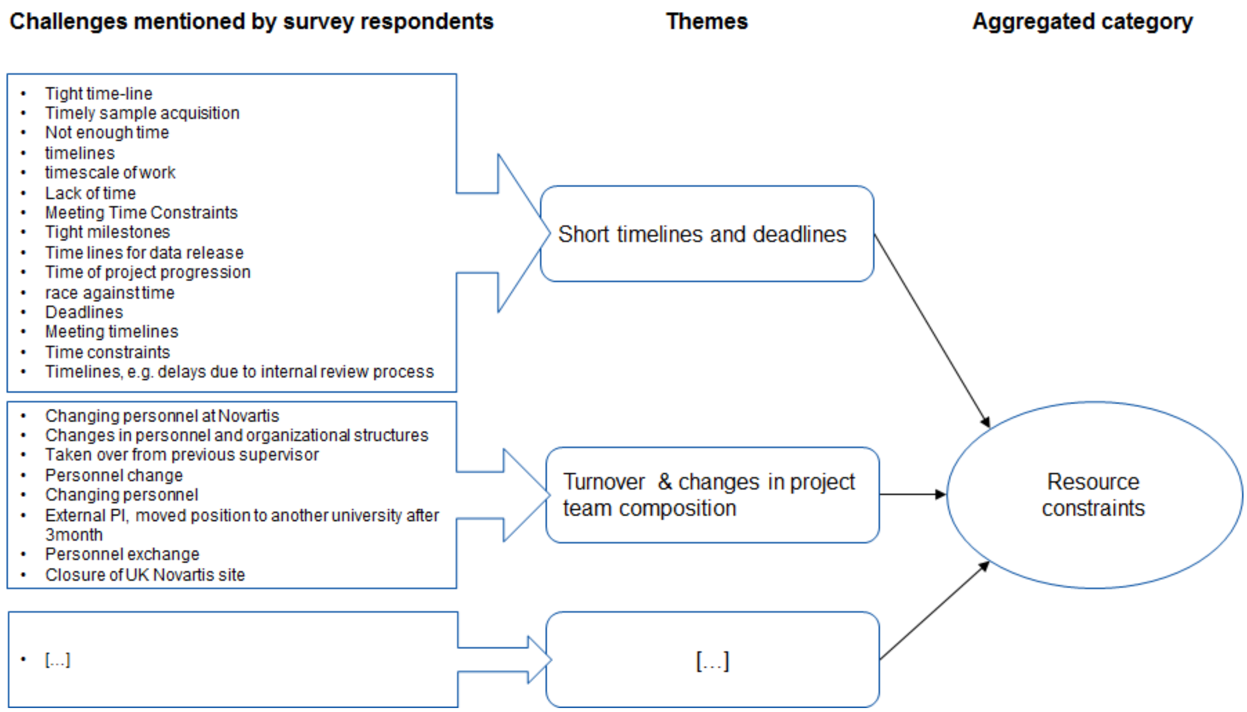

Figure S2 | **Illustration of the coding methodology.** The figure shows a subset of the coding scheme in the category "Resource constraints" to illustrate the approach. The left column contains the raw responses that were coded into themes. The centre column shows 2 out of 9 themes within the category "resource constraints". Afterwards, the themes were aggregated into categories, which constitute the final set of overarching categories in our analysis (right column, showing 1 out of 7 categories).

Three of the authors discussed this provisional coding scheme, analysed similarities and differences between the responses assigned to each theme and refined the coding scheme[5]. The authors also aggregated the themes to more abstract categories by grouping them according to their meaning (see Figure S2 for an illustration of this procedure). The coding scheme was refined and re-applied to the data. Five iterations of this procedure yielded the first version of the coding scheme. As figure S2 shows, our coding scheme involves two hierarchical levels. In the first level, responses that relate to the same topic, typically share a large fraction of the wording, or represent synonyms, are grouped into themes. In the second level, we interpreted these themes to group them into aggregated categories.

Third, we tested the reliability of the first version of the coding scheme with two additional research assistants, who are not part of the author team. Specifically, we tested the degree to which two different people would assign the text-based responses to the same category. We sent both research assistants the data, the list of the themes in the first version of the coding scheme (67 themes in total), a document detailing the coding procedure and a glossary of company-internal abbreviations and industry-specific terms that survey participants used. Both research assistants were tasked to independently assign the original data to the themes in the coding scheme. A comparison of the research assistants' application of the assignment showed a good agreement and therefore a high inter-coder reliability: both research assistants assigned 61% of the data (502 out 824 responses) to the same theme. Given that there was a large number of themes (67 themes in total in the first version of the coding) and comparing this agreement to the standards and recommendations for qualitative data analysis[11], these numbers suggest a good reliability of our coding scheme[5].

Fourth, we refined the coding scheme by comparing our own assignment of the responses into the different themes with the assignment by both research assistants. This refinement comprised the following steps:

I. Join themes that describe very similar challenges (4 themes joined into 2 themes) and refine the description of some themes (6 themes)

II. Keep all assignments between responses and themes where our initial assignment matches the assignment of **both** research assistants (409 responses, 50%). Drop those responses that were marked as ambiguous by **both** research assistants (7 responses, 1%)

III. Keep assignments where our initial assignment matches the assignment of **one** research assistant (193 responses, 23%), but drop responses where the discrepancies between the research assistants were large and indicated that the response could not be assigned unambiguously (22 responses, 3%)

IV. For all assignments where **none** of the research assistant's assignments matched with our original assignment: Discuss the disagreement with the research assistants and jointly agree on the preferable assignment. Select either the research assistants' assignment (90 responses, 11%) or the original response (74 responses, 9%) as the final assignment. Drop responses where no final agreement was found (29 responses, 4%).

Excluding the dropped responses, the refined coding scheme comprised 766 responses assigned to 65 themes. We dropped 44 responses that were assignment to the theme "No challenges", in which the respondents verbally indicated the absence of any challenges (e.g., by stating "no challenges"). Table S1 shows the compositions of the data set. In the final coding scheme, 722 responses were assigned into 64 themes, which were aggregated to 7 overarching categories.

Results. We computed the number of responses that were coded into each of the seven aggregated categories describing the different types of challenges as shown in Table S2. We ranked the challenges categories by the total number of responses. Furthermore, we computed the relative percentage of responses within each category in relation to the overall number of responses, separately for Novartis and academic respondents. The results are displayed in Figure 1 in the main article.

Table S2 | **Total number of responses categorized in the categories of challenges**[a]

Category	Novartis	Academic	Total	Rank
Resource constraints	101	66	167	1
Legal and administrative process complexity	84	53	137	2
Coordination challenges	76	60	136	3
Scientific challenges	60	67	127	4
Goal alignment challenges	44	10	54	5
Interpersonal challenges	36	16	52	6
Technological challenges	27	22	49	7
Total	428	294	722	

[a] The table lists the number of responses that we coded into the categories of challenges using the final coding scheme, shown separately for Novartis and academic respondents. The challenges are ranked using the total number of responses coded into each category.

References

1. Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **3,** 77–101 (2006).
2. EvaluatePharma. *World Preview 2017, Outlook to 2022*. (2017).
3. Novartis AG. *Novartis Annual Report*. (2017). Available at: https://www.novartis.com/sites/www.novartis.com/files/novartis-annual-report-2017-en.pdf
4. Nature Publishing Group. Nature Index 2017: Top 100 corporate institutions. (2017). Available at: https://www.natureindex.com/supplements/nature-index-2017-science-inc/tables/corporate-institutions.
5. Miles, M. B., Huberman, M. A. & Saldana, J. *Qualitative Data Analysis: A Methods Sourcebook*. 381 (SAGE Publications, 2014).

Supplementary Box 2 | **Data source and methodology for the quantitative analysis**

Research design. To test the extent to which the challenges identified in the qualitative analysis (see Supplementary Information 1 (box)) impact project success, we carried out a quantitative analysis. Our research design comprises two different questionnaires that were both sent to the project members of the collaborations in our sample. In our first questionnaire (the same used for identifying the challenges), we quantitatively measured a large number of factors that probe different project characteristics and team dynamics. To do so, we selected more than 40 different factors that are established concepts in management science. Our selection is based on the management literature on industry-academia collaboration, and interviews with seasoned experts in industry-academia collaboration in the pharmaceutical industry, technology transfer office professionals, and principal investigators at universities. Each factor was queried with one or more questions in the survey to reduce interpretation bias. Note that these factors were not chosen based on the categories of challenges, allowing the coding of responses into categories without prior bias (see Supplementary Information 1 (box) for information on the establishment of the categories of challenge).

The two questionnaires were spaced by about 6 months in a time-lag setup. This approach allows us to probe the factors related to the categories of challenges and project success separately in order to separate causes and effects. Unlike a cross-sectional research design that relies on only one questionnaire, a time-lag design reduces—albeit does not eliminate completely—the likelihood of a bias from reverse causality[1,2]. The detailed procedure of the data collection is described in the Supplementary Information 1 (box). Overall, we collected responses for questionnaire 1 from 669 project members in 187 projects (416 from NIBR project members, 253 from academic project members; overall response rate 85%) and responses for questionnaire 2 from 511 members in 168 projects (313 from NIBR project members, 198 from academic project members; overall response rate 76%).

Post-hoc, we associated each challenge category identified in the qualitative analysis with a quantitative factor exemplary for that category. The measurement of quantitative factors using a survey is based on research techniques established in management science[1,2]. In this approach, factors that describe both the characteristics and the dynamics of a project are measured using the assessment of all project team members. Querying and aggregating the perspectives of all team members helps to reduce the subjectivity bias. For each factor (e.g., the level of trust among project participants), the questionnaire included a set of questions that probe this factor on standardized Likert-scales (e.g., from 1 = “very low” to 5 = “very high”). The usage of multiple questions yields better reliability because it reduces bias due to different interpretations of the question[1].

To measure the outcome, we created a measure of project success based on similar measures in the management literature. Our conceptualization of project success measures the extent to which the project has achieved, or has made progress in achieving, the project's goals and is based on established measures[3,4]. As the intermediate outcomes of industry-academia collaborations are hard to measure and quantify[5], we rely on the assessment of project success by the project leaders from *both* organisations. The advantage of this approach is that it measures the performance of a project relative to the goals of the project that were jointly set by both project partners.

Quantitative data analysis. After the data collection had been concluded, we first carried out the inductive analysis (see Supplementary Box 1) to group the challenges reported by project participants into overarching categories. Afterwards, we associated each category with a factor from our first questionnaire that best represents this category. Table S3 provides an overview of all associations. For example, we relate the category "coordination challenges" from the inductive analysis to the factor "lack of coordination" that we measure with a previously selected measure of the degree of coordination (selected independent of inductive analysis). These factors serve as proxies to describe the extent to which challenges from that category are present in a project. Although these factors can by design not measure all aspects of each category of challenges, they represent factors that were named frequently in the qualitative analysis. The advantage of using pre-defined, concise factors from the previous literature is that established quantitative scales are available, which allow a reliable measurement of these factors.

As Table S3 shows, all but one of the seven factors that we measured are quantitative in nature. That is, participants responded to the questions in the questionnaire on quantitative scale. Since the contract type is a categorical variable, we grouped the available contract options into low (e.g., template-based, such as a material transfer agreement or a confidentiality agreement) and high (e.g., customized, such as a research agreement) complexity contracts depending on the effort required to set up such a contract based on the experience of senior managers at Novartis. In the subsequent analysis, we aggregated all responses within a project that measure the same variable.

To assess project success, we used the responses from the project leaders from both organizations in the second questionnaire (259 responses in total, 144 from Novartis project leaders and 115 from academic project leaders). We further restricted our analysis to projects for which we had received at least three responses and at least one response from each organisation (120 projects total) to reduce the potential subjectivity bias in the responses. To ensure the reliability of the outcome measure *project success*, we examined intra-class correlation coefficients (ICCs). The results showed a good agreement between the two project leaders that assessed the success of a project (the ICC(1) value for the overall project success measure was 0.54, indicating a high agreement between the different ratings for the same project[6]). Therefore, we aggregated the project leaders' rating of the outcome variable *project success* for each project.

Table S3 | **Measures quantifying factors related to the categories of challenges**

Category of challenges	**Related factor: description**	**Measure**
Resource constraints	*Lack of resources:* The home organisation provides little flexibility to pursue creative ideas and to spend time on projects except pipeline projects.	Quantitative factor based on 1 multiple choice question (Likert scale) from an established measure of the organisational innovation climate[7]
Legal & administrative process complexity	*Legal contract complexity:* The underlying legal contract is not based on standardized or template-based contracts, but is customized and typically requires negotiation.	Binary measure: 0=low-complexity contract 1= high-complexity contract
Coordination challenges	*Lack of coordination:* Problems in coordinating tasks and activities with the team at the partnering organisation.	Quantitative factor based on 5 multiple choice questions (Likert scale) from an established measure in the multi-team literature[8]
Scientific challenges	*Scientific uncertainty:* The scientific phenomena are not well understood, there is much trial and error, and cause-and-effect relationships are largely unknown.	Quantitative factor aggregated from 5 multiple choice questions (Likert scale), adapted from an existing measure of uncertainty in product development projects[9]
Goal alignment challenges	*Goal discrepancy:* There is a low overlap between the goals of the two partnering organisations.	Quantitative factor based on 1 question using a graphical measure depicting various configurations of goal alignment, adapted from an established measure[10]
Interpersonal challenges	*Lack of trust:* Participants do not trust participants from the other organisation.	Quantitative factor based on 4 multiple choice questions (Likert scale) from an established measure in the alliances literature[11]
Technological challenges	*Employing explorative technology:* The project involves technology, methods and compounds that are based on fundamentally new concepts or principles.	Quantitative factor based on 2 questions using a continuous quantitative scale 1-100 from an established measure in the R&D teams literature[12]

[a] The table shows the selection of seven factors that relate to the categories of challenges identified from questionnaire 1 and describes the employed measure. Most measures have been adapted from the literature to fit the context of industry-academia collaborations in the pharmaceutical industry.

To infer the relationship between the challenge-related factors and project success, we stratified all projects in our sample into two groups for each factor, sorted into projects scoring (1) high or (2) low in that factor. We considered all projects with a score higher than the mean in that factor as “high”, and below the mean as “low”. We then calculated the average level of project success separately for (1) the projects that face a high level of challenges and (2) the projects that face a low level of challenges in the corresponding category. Finally, we employed an ANOVA (analysis of variance) statistical test to infer whether the groups scoring high and low in a specific factor also differ significantly in their average rating of project success. Since the outcome measure *project success* was only fully reported by 110 teams, this analysis includes 110 projects in our sample.

Table S4 | **The impact of factors related to challenges on project success**[a]

Factor	**Average rating of project success**		**Differences between projects scoring low and high in factor (ANOVA)**
	Projects scoring low in factor (below average)	**Projects scoring high in factor (above average)**	
Lack of resources	4.12 (N=51)	3.79 (N=49)	Significant at p<0.02
Legal contract complexity	3.84 (N=58)	4.08 (N=47)	Significant at p<0.09
Lack of coordination	4.21 (N=57)	3.64 (N=52)	Significant at p<0.001
Scientific uncertainty	4.08 (N=55)	3.81 (N=55)	Significant at p<0.04
Goal discrepancy	4.14 (N=58)	3.73 (N=52)	Significant at p<0.002
Lack of trust	4.04 (N=55)	3.85 (N=55)	Not significant
Employing explorative technology	3.82 (N=54)	4.07 (N=54)	Significant at p<0.06

[a] The table shows the average rating of project success for (i) projects that score low in a specific factor and (ii) projects that score high in that factor. The overall sample of 110 projects is split into these two groups for each factor separately (the numbers do not add up to 110 in four cases due to missing data). The column on the right shows the results from analyses of variances (ANOVAs), indicating whether there is a significant difference in the average rating of project success for projects that score low and for those that score high in the specific factor.

Results. Table S4 presents the results, which are plotted in Figure 2. The results indicate that a higher manifestation of five out of the seven factors examined is associated with a lower average level of project success. For example, projects scoring above average with respect to coordination problems (N=52) exhibited an average rating of project success of 3.64 out of 5. This value is significantly lower than the average rating of projects success of 4.21 for projects that scored below average in coordination problems (an analysis of variance (ANOVA) showed that the difference is significant at p<0.001). We found similar results for four other factors (i.e., lack of resources, scientific uncertainty, goal discrepancy, and lack of trust). The strongest impact on project success arises from of a lack of coordination.

References

1. Wolf, C., Joye, D., Smith, T. & Fu, Y.-C. *The SAGE Handbook of Survey Methodology*. (SAGE Publications, 2016).
2. Easterby-Smith, M., Thorpe, R. & Jackson, P. R. *Management Research*. (SAGE Publications, 2012).
3. Bryde, D. Perceptions of the impact of project sponsorship practices on project success. *Int. J. Project Manage.* **26,** 800–809 (2008).
4. Lewis, M. W., Welsh, M. A., Dehler, G. E. & Green, S. G. Product Development Tensions: Exploring Contrasting Styles of Project Management. *Acad. Manage. J.* **45,** 546–564 (2002).
5. Perkmann, M., Neely, A. & Walsh, K. How should firms evaluate success in university-industry alliances? A performance measurement system. *R&D Management* **41,** 202–216 (2011).
6. Bliese, P. D. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. in *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (eds. Klein, K. J. & Kozlowski, S. W. J.) 349–381 (Jossey-Bass, 2000).
7. Scott, S. G. & Bruce, R. A. Determinants of innovative behavior: A path model of individual innovation in the workplace. *Acad. Manage. J.* **37,** 580–607 (1994).
8. Hoegl, M., Weinkauf, K. & Gemuenden, H. G. Interteam Coordination, Project Commitment, and Teamwork in Multiteam R&D Projects: A Longitudinal Study. *Organization Science* **15,** 38–55 (2004).
9. Song, M. & Montoya-weiss, M. M. Technological the Effect of Perceived Uncertainty on Japanese New Product Development. *Acad. Manage. J.* **44,** 61–80 (2001).
10. Bartel, C. A. Social Comparisons in Boundary-Spanning Work: Effects of Community Outreach on Members' Organizational Identity and Identification. *Adm. Sci. Q.* **46,** 379–413 (2001).
11. Poppo, L., Zhou, K. Z. & Ryu, S. Alternative Origins to Interorganizational Trust: An Interdependence Perspective on the Shadow of the Past and the Shadow of the Future. *Organization Science* **19,** 39–55 (2008).
12. Miron-Spektor, E., Erez, M. & Naveh, E. The effect of conformist and attentive-to-detail members on team innovation: Reconciling the innovation paradox. *Acad. Manage. J.* **54,** 740–760 (2011).