DIW Weekly Report 19 2019

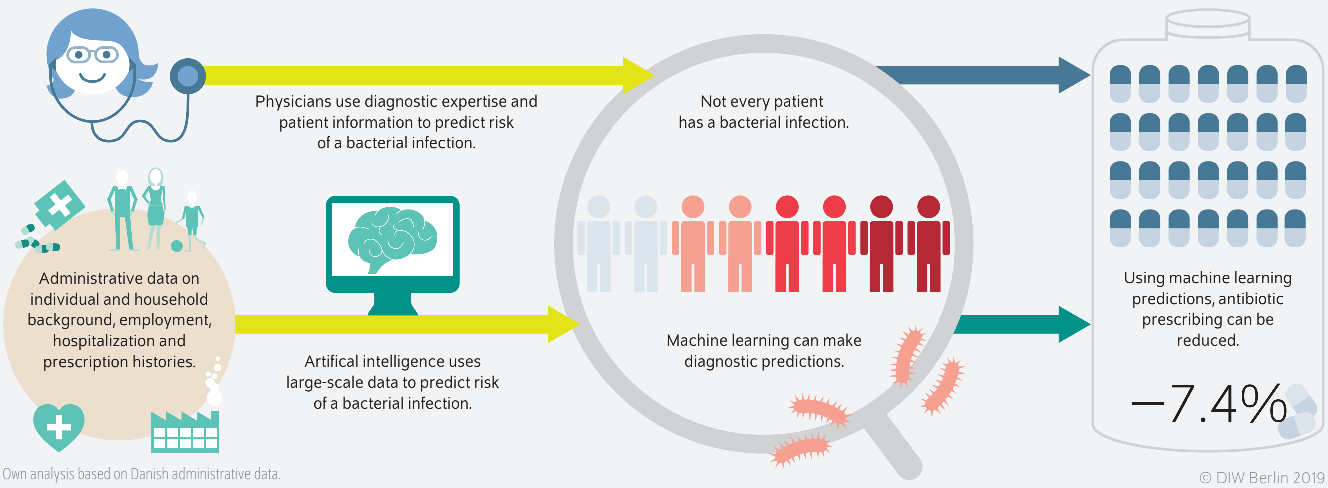
AT A GLANCE

Artificial intelligence and big data can help contain resistance to antibiotics

By Michael A. Ribers and Hannes Ullrich

- The subject of this study is whether big data and machine learning can improve physician prescription practices
- Machine learning and economic modeling are applied to comprehensive administrative and medical data
- Rules for prescribing are developed as a way of reducing the number of incorrect prescriptions of antibiotics
- The total number of antibiotic prescriptions for urinary tract infections can be reduced by 7.42 percent without reducing the number of treated bacterial infections by applying the recommended policy measures
- The potential of this method requires a large amount of digitalized patient-level information

Machine learning predictions can reduce antibiotic prescribing for suspected urinary tract infections by more than 7 percent



Own analysis based on Danish administrative data.

FROM THE AUTHORS

"Our results show the potential of data-based predictions using machine learning methods for containing resistence to antibiotics. Only the adequate prescription of antibiotics allows maintaining their therapeutic effectiveness."
— Hannes Ullrich, author of the study —

MEDIA

Audio Interview with Hannes Ullrich (in German)
www.diw.de/mediathek

Artificial intelligence and big data can help contain resistance to antibiotics

By Michael A. Ribers and Hannes Ullrich

ABSTRACT

Improving physicians' prescription practices is a primary strategy for countering the rise in resistance to antibiotics. This would prevent physicians from incorrectly prescribing antibiotics, one of the main causes of antibiotic resistance. The increasing availability of medical data and methods of machine learning provide an opportunity to generate instant diagnoses. In the present study, the example of urinary tract infections in Denmark is used to demonstrate how data-based predictions can improve clinical practice in the face of increasing antibiotic resistance. For this purpose, comprehensive administrative and medical data, in combination with machine learning methods and economic modeling, were used to develop rules for prescribing antibiotics. The total number of prescriptions could be reduced by 7.42 percent by applying the recommended policy measures without reducing the number of treated bacterial infections. This demonstrates the great potential of this method. However, in Germany this potential cannot be tapped until more information is digitized. The information that must be supplied to the IT systems in physicians' practices and hospitals is often collected and saved by decentralized institutions; linking it is key.

In the 1940s, the development of penicillin for use in treatment revolutionized medicine and enabled new medical procedures. Penicillin has saved millions of lives since then. The increasing use of antibiotics for both medical and non-medical purposes—in the agriculture industry, for example—has given rise to the problem of growing resistance to antibiotics over the past few decades. Simple infections caused by multi resistant bacteria—including widespread *Escherichia coli*, *Klebsiella pneumoniae*, and *Staphylococcus aureus*—are becoming life-threatening risks again. Every year, around 700,000 people worldwide die from infections caused by resistant bacteria. The World Health Organization considers antibiotic resistance one of the major health policy challenges of our era. The annual cost linked to resistant pathogens has been predicted to reach 100 billion dollars by 2050.[1] A lack of financial incentives for developing new antibiotics is the main problem for the health care sector, but this can be addressed by implementing supply-side policy measures such as subsidies or exceptions within the patent system.[2] Resistance has developed to all newly developed antibiotics, however, causing their therapeutic value to diminish over time. This puts the focus of health care policy on physicians' prescription practices, which play a leading role in preserving the therapeutic benefit of existing medications.[3]

The study presented here was funded by the European Research Council and carried out at the German Institute for Economic Research (DIW Berlin). Using suspected urinary tract infections in Denmark as an example, it shows how policy measures based on machine learning and comprehensive individual data can reduce the number of unnecessary prescriptions and increase the number of expedient prescriptions.[4] In particular, it shows that machine learning predictions can differentiate among bacterial and non-bacterial infections with such high certainty that a reduction

1 Jim O'Neill, Tackling drug-resistant infections globally: final report and recommendations (2016) (available online).

2 See Ramanan Laxminarayan et al., "Antibiotic resistance—the need for global solutions," *The Lancet Infectious Diseases*, 13(12) (2013): 1057-1098.

3 See Eili Y. Klein et al., "Global increase and geographic convergence in antibiotic consumption between 2000 and 2015," *Proceedings of the National Academy of Sciences*, 115(15) (2018): E3463-E3470.

4 See Michael A. Ribers and Hannes Ullrich, "Battling Antibiotic Resistance: Can Machine Learning Improve Prescribing?," *DIW Discussion Paper* No. 1803 (2019).

DOI: https://doi.org/10.18723/diw_dwr:2019-19-1

in antibiotic prescribing of around 7.42 percent is possible without reducing the number of treated bacterial infections. Since this finding is based on the redistribution of prescriptions, and some patients will have prescriptions delayed, the associated risks are also discussed. Finally, the authors present their reflections on the challenges that comparable measures in Germany would face.

Antibiotic prescribing promotes the development of resistance

Alongside a lack of hygiene and the use of antibiotics in agriculture, the human ingestion of antibiotics is a principle reason why bacteria become resistant to one or more antibiotic molecules. The more bacteria are exposed to antibiotics the more become resistant; insofar as they have survived their encounter with these substances. Both unnecessary over-prescribing that exposes bacteria to antibiotics and under-prescribing that enables possible resistant bacteria to survive and evolve support resistance.

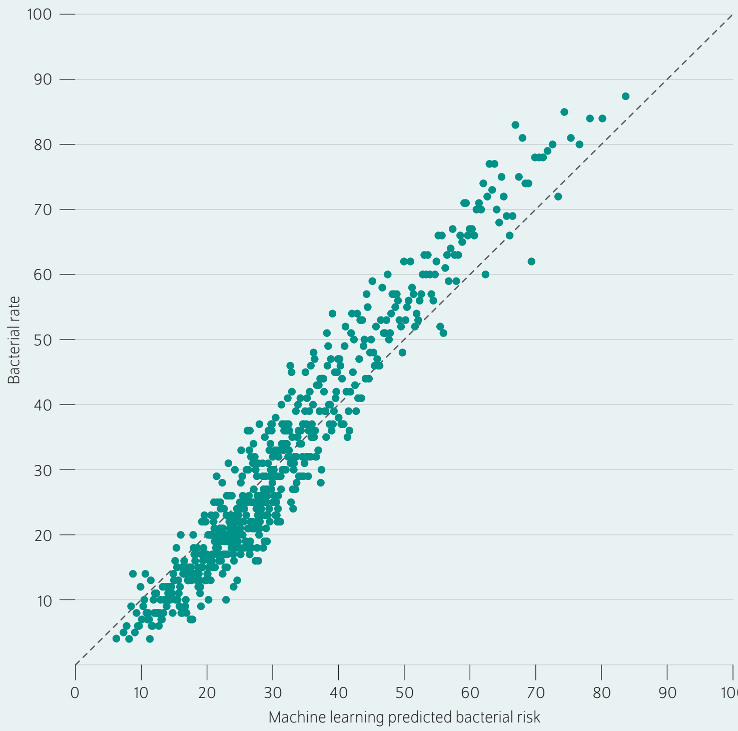
In a typical general practice treatment situation in which a patient presents the symptoms of an infection for the first time, physicians must solve two problems. First, they have to create a diagnosis that contains an assessment of the possible cause. Second, they must weigh the expected positive effect of treatment against the expected development of resistance. The currently available diagnostic tools in general practitioners' practices result in uncertain diagnoses, particularly for first examinations.[5] For urinary tract infections they can request detailed microbiological laboratory tests that will identify possible pathogens and create information on their resistance profiles. But this information is usually only available after a several-day delay, often equal to the duration of a treatment unit with an antibiotic. Patients usually require immediate treatment. As a consequence, decision making under uncertainty on the part of physicians leads to over- and under-prescribing.

Developing rapid diagnostic tools is described as an important objective in restricting unnecessary prescriptions.[6] Data-based procedures are already being allocated a potential role.[7] However, instant detailed diagnoses that provide information on specific bacteria and the efficacy of individual antibiotics will remain a challenge in the foreseeable future. For the large majority of prescription decisions, it would be helpful to have information on the probability of bacterial infection at hand. It would be the basis for informed decisions on whether or not treatment with antibiotics could be delayed until detailed test results are available or should be initiated immediately.

Figure 1

Bacterial rate by predicted bacterial risk in percent

Each sphere represents 100 observations sorted by risk



Source: Own analysis based on Danish administrative data.

The proximity of spheres to the diagonal show the high prediction quality.

Machine learning and big data can improve prescription practices

Physicians in principle possess a great deal of background information on patients that can help generate an initial diagnosis. They know whether the person on the other side of the desk has most likely been exposed to bacterial pathogens (because he/she works in livestock farming and not work in construction, for example). They can also collect information about individual treatment, diagnostics histories, and much more. Because this information is high-dimensional and potentially vast, it takes a lot of effort to analyze. Evaluating it requires extraordinary amounts of information processing capacity, expertise, and time. Artificial intelligence delivers tools that can reduce this complexity.

Machine learning is artificial intelligence's most important tool. Its strength is solving frequently occurring, clearly definable problems about which a great amount of information in the form of digitized data is available (see Box 1). Data availability and technological progress have created the technological and economic ability to make meaningful predictions based on a wealth of information that is easy to interpret.[8]

5 See Angela M. Caliendo et al., "Better tests, better care: improved diagnostics for infectious diseases," *Clinical Infectious Diseases*, 57 (suppl_3) (2013): 139-170.

6 Jim O'Neill, "Tackling drug-resistant infections globally."

7 Jim O'Neill, "Tackling drug-resistant infections globally," 37.

8 Ajay Agrawal, Joshua Gans, and Avi Goldfarb, "Prediction machines: the simple economics of artificial intelligence," *Harvard Business Press*, (2018).

Box 1

Machine learning for policy making

Machine learning makes predictions solely on the basis of correlations between observed outcomes and predictor variables. Sometimes variables that are important for predicting a result are found—in the present context, for example, the time and type of the last antibiotics prescription. These variables are key contributors to the quality of the prediction. But they do not permit conclusions about causality. The decision about a previous treatment is unlikely to be the cause of a current bacterial infection. It is much more likely that a factor not observed in the data, such as high consumption of contaminated food, led to both the previous treatment decision and the current infection. Such differentiation is irrelevant for prediction quality because past treatment decisions capture, at least partially, this information.

Economists have traditionally examined causal relationships in order to quantify the effects of a policy measure on (market) outcomes. Through the availability of detailed high-frequency data, the development of algorithms, and the increase in computing capacity new opportunities arise to focus on policy measures that contain prediction problems.[1] As in the present study, this could be to estimate the risk of a bacterial infection. In the U.S., the use of algorithms and data to estimate the recidivism rate of defendants awaiting bail decisions received much attention.[2] Beyond the valid criticism, it was also possible to show the potential for lowering crime rates and lower levels of discrimination against disadvantaged groups.[3] The careful context-specific evaluation of new data-based policy interventions, therefore, appears to be a highly promising field for further research and public discussion. Alongside the quantification of the potential of possible applications, it is essential to consider the ethical dimensions of implementation, with relation to discrimination and fairness, for example.[4] The relevant comparison along such dimensions is between data-based predictions and observed human action, which in itself is not free from discrimination.

1 See Jon Kleinberg et al., "Prediction policy problems," *American Economic Review Papers and Proceedings*, 105(5) (2015): 491-495; and Helen Margetts and Cosmina Dorobantu, "Rethink government with AI," *Nature*, 568 (2019): 163-165.

2 See Julia Angwin et al., "Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks," ProPublica, May 23, 2016 (available online, accessed on April 9, 2019); and Sam Corbett-Davies et al., "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear," *The Washington Post*, October 17, 2016 (available online, accessed on April 9, 2019).

3 Jon Kleinberg et al., "Human decisions and machine predictions," *The Quarterly Journal of Economics*, 133(1) (2018): 237-293.

4 See Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, "Inherent trade-offs in the fair determination of risk scores," *arXiv* (2016) (available online).

In medicine, the perfect example is image recognition in radiology or ophthalmology, where the high case frequency, the digitization of images and classifications, and the relative simplicity of the problem make it suitable for everyday use.[9]

The study presented here used an approach that enabled the comparison of machine learning predictions with physicians' prescription practices. It provides the basis for data-based prescribing rules that could improve prescription practices. In detail: the predicted bacterial risks for individual patients were used in order to prevent physicians from making instant prescriptions to patients with low predicted risk and enable patients with high predicted risk to obtain prescriptions. To do this, the laboratory test results of patients with presumed urinary tract infections from the largest medical laboratory in Denmark were anonymized and linked to rich administrative data. The data included complete prescription and treatment histories, past hospital stays, laboratory results, and demographics such as age, gender, profession, municipality of residential address, household size and type, marital status, and more. As a result, a data set with 95,594 observations and 1,266 variables collected over three years was available for analysis. The underlying administrative data have been collected across Denmark, saved centrally, anonymized, and made available for research purposes for some time. The basic prerequisite for the application presented here was the linkage of this wealth of information across individuals and time.

To predict bacterial causes, the authors used a random forest based on regression trees (see Box 2).[10] The quality of the predictions is shown based on the average rate of observed bacterial infections for patients with various values of the predicted risk of a bacterial infection (see Figure 1). Each mapped point stands for 100 test observations for which the average rate was calculated. Ideally, all points are on the diagonals. The proximity of the points to the diagonals is an indication of the high prediction quality achieved by the random forest based on the administrative data used.

However, the pure quality of the prediction is not decisive but rather its comparison with physicians' prescribing decisions. If physicians have better information than that contained in the data used by the algorithms, or have context-specific expertise, it would not be surprising if the predicted risks did not contain additional information. Indeed, the findings showed that non-observable information and expertise lead to patients with a prescription for antibiotics (*treated*) having a higher rate of bacterial infection—given the predicted risk—than patients without prescriptions (*untreated*)

9 See Ziad Obermeyer and Ezekiel J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England Journal of Medicine*, 375(13) (2016): 1216.

10 See Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning: data mining, inference, and prediction* (New York: Springer Series in Statistics, 2009).

(see Figure 2). Diagnostic methods that provide physicians with some, imperfect same-day information on the presence of a bacterial infection can partly explain these results as the same-day diagnostics were not available to the algorithm. Detailed observation in the study showed clear differences in the ability of physicians to dispense prescriptions in accordance with predicted risk. This observation indicates differences in available diagnostics methods, expertise, and prescription preferences.

It is evident that a relevant number of prescriptions is dispensed to patients with low predicted risk (see Figure 3, red triangles). And a relevant number of patients with high predicted risk do not receive a prescription (see Figure 2, light-green dots). The proposed measure for reducing over-prescription examined in the study focuses on these findings. Subject to the condition that the total number of treated bacterial infections remains constant, prescriptions for people with low risk can be redistributed to those with high risk. This could reduce the total number of prescriptions for antibiotics for urinary tract infections by 7.42 percent. Alternatively, for a constant number of total prescriptions, an increase in treatments of bacterial infections with antibiotics could be increased by 6.38 percent.

Since this is a case of redistribution, a justified concern is that prescriptions could be delayed for patients whose health risk is exceptionally high due to the presence of a bacterial infection. For example, untreated infections can lead to complications in the case of pregnancy, which accounts for approximately one-third of treatment situations. In order to examine this issue, the same redistribution was carried out subject to the condition that the physician's decision for these women must remain unchanged. At 6.81 percent, the possible reduction in instant prescriptions is somewhat lower but, in the analysis, the rates are statistically not distinguishable.

Figure 2

Bacterial rate by predicted bacterial risk in percent conditional on prescription decisions

Each sphere and each triangle represents 100 observations

Source: Own analysis based on Danish administrative data.

Red triangles and light-green spheres show areas in which machine learning can improve antibiotic prescribing.

Conclusion: machine learning methods have great potential but still face challenges

The results show the potential for data-based predictions using machine learning methods. The available mass and complexity of data can be reduced to simple useful information and in this way contribute to solving one of the major challenges of today's health care system: the efficient use of antibiotics given the trend toward sharply rising resistance. Due to their great potential, the methods could be tested in practice with physicians—even though their implementation in the context of general practitioners would require considering (presumably manageable) risks for individual patients. At the same time, another endeavor shows additional potential: combining comprehensive background information with case-specific information that physicians can collect during treatment. For example, symptoms and behavior-based risk factors could be included in bacterial risk prediction. In Denmark, the technical prerequisites for collecting such data already exist in principle, as physicians' practices are digitized and linked to centralized servers on which physicians can retrieve patient data.

The findings of the study in the Danish context are based on satisfying three key conditions. They provide information on the challenges that would have to be overcome to consider comparable applications in Germany.

First, the availability of digitized, standardized, linked, and centrally accessible data is key to the success of machine learning. To achieve systematic digitalization and standardization, processes are required that generate data as a side product of services in the health care system or that systematically store information for specific applications. Linkage and central availability are two particular advantages of the centralized Danish health care system. They minimize the interfaces and heterogeneity of data sources and facilitate large-scale data collection. In Germany, private actors such as hospitals and insurance companies are making decentralized efforts to collect and link medical and background data.[11] Their decentralized nature makes it extremely difficult to use the data they generate to design and implement

11 For example, see Hauke Hohensee, "Was bringt smarte Technik für die Gesundheit?" (2019) (in German; available online, accessed on April 9, 2019).

Box 2

Machine learning example: the random forest

Easy to implement and very flexible, random forests are among the leading standard algorithms of machine learning due to their high prediction quality.

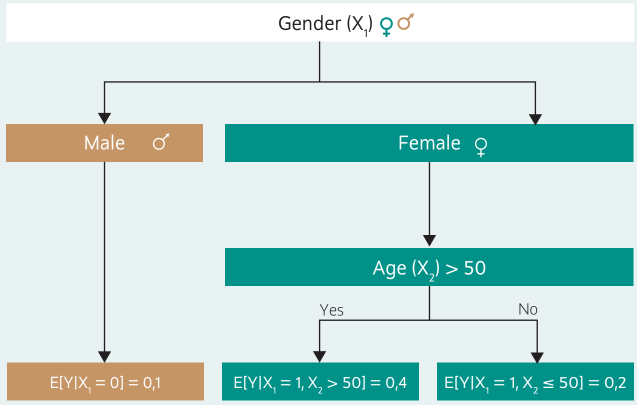
A random forest is a collection of classification or regression trees. A regression tree generates the conditional expectation $E[Y|X]$ of the outcome variable Y based on a collection of M observed variables X. If Y is binary, a regression tree predicts the conditional probability that a bacterial cause is present in the example at hand. A simplified example with two variables illustrates how the algorithm searches for the best grouping by X, variable by variable, in order to maximize the prediction quality of $E[Y|X]$ (see Figure 3). This process is called "training" and is based on training data.

Regression trees have low bias, meaning that individual observations can be predicted very well. However, they have large prediction errors out of sample, that is for data that differs from the training data. In order to reduce this variance, extra steps are carried out in a random forest.

1. For each tree, a random sample of observations from the training data set are used.
2. At each node in the tree, $m<M$ variables are randomly drawn for which classifications are sought to maximize the quality of the prediction.
3. The prediction is calculated for each observation as the average of all conditional expectations obtained from all trees.

Figure 3

Simple regression tree



A regression tree searches the best grouping by X in order to maximize the prediction quality of E[Y|X].

policy measures. The problems encountered as digital health insurance cards were being implemented indicate just how challenging it is to set up centralized solutions within the German institutional context. Given the health care policy challenges, and unlike many decentralized health care services, the non-bureaucratic implementation of a central data pool—which at least ensures system interconnectivity—would create myriad conditions for improving health care provision.

Second, the positive results in the context of prescribing antibiotics were achieved by reducing the problem to simple binary variables that generated a recommendation for a yes-no decision. As described above, we believe that simple, conventional problems of prediction for which an enormous pool of relevant information is available have the greatest potential. Ambitious systems such as IBM Watson have failed to solve more complex problems, clearly indicating that as of today, data availability coupled with efficient algorithms and high computing capacity do not automatically lead to intelligent predictions.[12] Due to problem-specific differences in data collection, the information content of the collected data, and the significance of non-digitized information that algorithms must leave out, data-based systems must be implemented and evaluated within specific contexts. Consequently, we can conclude that artificial intelligence is not yet suitable for general prediction across a wide spectrum of tasks. Its strength lies in clearly definable problems for which the relevant data can be collected in sufficient magnitude. In the short term, it would therefore pay to look for simple problems that can be automated and tap the potential there.

An important third challenge would be to make fair substantiated comparisons between data-based decisions and human decisions based on expert knowledge. For such comparisons, analytical approaches must be used that consider the data-generating decision processes that can lead to distorted data or selected samples.[13] Therefore, Germany must invest even more in educating excellent, quantitative, and multidisciplinary experts. Due to their focus on data generation and human decision-making, the empirical economic sciences could make a solid contribution, but other empirical

12 Eliza Strickland, "IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care," *IEEE Spectrum*, 56(4) (2019): 24-31.

13 See Janet M. Currie and W. Bentley MacLeod, "Understanding physician decision making: the case of depression," *NBER Working Paper*, no. w24955 (2018); Kleinberg et al., "Prediction policy problems"; and Himabindu Lakkaraju et al., "The selective labels problem: evaluating algorithmic predictions in the presence of unobservables," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2017): 275-284.

disciplines could also provide helpful approaches.[14] In strategies for promoting artificial intelligence, such expertise would be decisive for its successful implementation in a wide range of policy applications above and beyond technical investment.

In order to develop the potential of machine learning methods and promote their social acceptance at the same time, experimental trials in collaboration with experts in the field and accompanying evaluation are essential. Machine learning may have the potential to contribute to considerable social progress, but these significant challenges indicate that its full impact can only be achieved step by step. In this regard, society will have to wait for the great leap forward.

14 Susan Athey, "The impact of machine learning on economics," in "The economics of artificial intelligence: an agenda," eds. Ajay Agrawal, Joshua Gans, and Avi Goldfarb (Chicago: University of Chicago Press, 2018).

Michael A. Ribers is Research Associate at DIW Berlin and Postdoctoral Researcher at the University of Copenhagen | mribers@diw.de

Hannes Ullrich is Research Associate at DIW Berlin and Associate Professor at the University of Copenhagen | hullrich@diw.de

JEL: C10; C55; I11; I18; L38; O38; Q28

Keywords: Antibiotic prescribing; prediction policy; machine learning; expert decision-making

LEGAL AND EDITORIAL DETAILS

DIW BERLIN

DIW Berlin — Deutsches Institut für Wirtschaftsforschung e. V.
Mohrenstraße 58, 10117 Berlin
www.diw.de
Phone: +49 30 897 89–0 Fax: –200
Volume 9 May 7, 2019

Sale and distribution
DIW Berlin Leserservice, Postfach 74, 77649 Offenburg
leserservice@diw.de
Phone: +49 1806 14 00 50 25 (20 cents per phone call)

Layout
Roman Wilhelm, DIW Berlin

Cover design
© imageBROKER / Steffen Diemer

Composition
Satz-Rechen-Zentrum Hartmann + Heenemann GmbH & Co. KG, Berlin

ISSN 2568-7697