Article

S|S|S

Social Studies of Science
2019, Vol. 49(4) 476–502

DOI: 10.1177/0306312719862049
journals.sagepub.com/home/sss
SAGE

Following the algorithm: How epidemiological risk-scores do accountability

Katrin Amelang
Department of Anthropology and Cultural Research, University of Bremen, Bremen, Germany

Susanne Bauer
Centre for Technology, Innovation and Culture (TIK), University of Oslo, Oslo, Norway

Abstract

Epidemiological risk scores are calculative devices that mediate and enact versions of accountability in public health and preventive medicine. This article focuses on practices of accountability by following a cardiovascular risk score widely used in medical counselling in Germany. We follow the risk score in the making, in action, and in circulation to explore how the score performs in doctor-patient relations, how it recombines epidemiological results, and how it shapes knowledge production and healthcare provision. In this way, we follow the risk score's various trajectories – from its development at the intersection of epidemiology, general medicine and software engineering, to its usage in general practitioners' offices, and its validation infrastructures. Exploring the translations from population to individual and back that are at work in the risk score and in the primary prevention of cardiovascular disease, we examine how versions and distributions of accountability are invoked and practiced as the score is developed and put to use. The case of a simple risk score used in everyday counselling brings into relief some key shifts in configurations of accountability with emerging versions of 'health by the algorithm'. While there is an increasing authority of algorithmic tools in the fabric of clinical encounters, risk scores are interwoven with local specificities of the healthcare system and continue to be in the making.

Keywords

accountability, algorithms, calculative devices, cardiovascular prevention practices, doctor–patient interaction, risk scores

Correspondence to:
Susanne Bauer, Centre for Technology, Innovation and Culture (TIK), University of Oslo, P.O. Box 1108, Blindern, 0317 Oslo, Norway.
Email: susanne.bauer@tik.uio.no

Introduction

This article is about emerging versions of accountability in practices that use epidemiological risk scores to guide decisions about disease prevention. Rather than take accountability as a given, we approach accountability as a host of relations that are negotiated and enacted through calculative practices. We explore how one device – a cardiovascular risk score widely used in medical counselling in Germany – configures accountability and contributes to the emergence of a specific mode of 'epidemiological accountability' (Reubi, 2018). Following the case of a risk score through its 'social life', we argue that its modes of implementing preventive medicine enrol individuals into preventive individual risk-factor optimizing measures, which emerge at the population level in observational epidemiology, thus mobilizing citizens – pre-patients – as entangled 'public health subjects'.

In contrast to many ideas about public health as a decisively 'social project' (Kickbusch, 1986), in West Germany epidemiology was mostly taken up as a biomedical model enriched by individual-level risk factors rather than a community health perspective (Niewöhner et al., 2011: 729–730; see also Madarász, 2010).[1] While medical practice stressed the individual level, population-level thinking was foundational in the epistemic infrastructure of study designs and biostatistical techniques. For cardiovascular epidemiology and cohort studies of chronic disease, the seminal study is the Framingham Heart Study in the US, which not only established the cohort model but also coined the term 'risk factor' (Aronowitz, 2011). This approach to epidemiology continues to shape international research on cardiovascular disease, despite criticisms that the concept reifies the individualization of health by treating 'lifestyle' risk factors through pharmaceuticals while also facilitating the rise of surveillance medicine (Armstrong, 1995, 2014; Dumit, 2012; Greene, 2007). Other scholars debate the ways in which clinical trials and risk-factor epidemiology shape how matters of health are framed, categorized and accounted for (Aronowitz, 2009; Holmberg et al., 2013; Marks, 1997; Shim, 2014; Will and Moreira, 2010). Our article explores the entanglements and translations of practices with one specific epidemiological risk score used in preventive medicine. We discuss practices associated with the risk score and the negotiations taking place as the score enters medical routines. In this way, we examine how risk scores enact accountability in practice, in terms of who is held accountable for what and in what ways, and how they configure and realign accountabilities in health care and public health.

The rise of evidence-based medicine in clinical decision-making (Sackett et al., 1996; Weisz, 2005) is associated with the quest for 'evidence-based public health'. One response to this quest has been the development of epidemiological risk scores, especially cardiovascular risk scores designed to support decision-making and preventive counselling. Risk scores combine several interacting risk factors to estimate the overall risk of cardiovascular disease of an individual. Internationally known examples are the Framingham Risk Score (D'Agostino et al., 2008; Wilson et al., 1998), QRISK2 (Hippisley-Cox et al., 2007), the FINRISK Calculator (Vartiainen et al., 2016), and the ESC-Score (Conroy et al., 2003); in Germany, SCORE Germany (Keil et al., 2005) and PROCAM (Assmann et al., 2002). As 'calculative devices' (Callon and Muniesa, 2005), these risk scores compute an individual estimate based on multivariate statistical models

Table 1. The six steps of an arriba consultation, according to arriba booklets.

1	Define task together	a
2	Subjective risk (patient's ideas and concerns)	r
3	Objective risk (calculation)	r
4	Informing about options for intervention	i
5	Joint evaluation of options	b
6	Agreement on action to follow	a

and findings from large epidemiological observation studies. For instance, cohort data sets comprise individual-level information for defined populations and are used as the foundation for the algorithm that computes a risk figure based on a set of variables known for this person. Thus, risk scoring enrols the individual into a population-level calculus. The use of risk scores to compute individualized probabilities of disease are key elements of 'evidence-based prevention' and 'personalized prevention', which are an offshoot from 'evidence-based medicine' and 'personalized medicine' (Hoeyer, 2019; Prainsack, 2014; Tutton, 2012).

Taking the risk score 'arriba', a counselling tool and a risk-predicting algorithm currently in use in general practitioners' offices in Germany, as a case study, we examine translations from population to individual and back in the primary prevention of cardiovascular disease. Our article builds on and extends Science and Technology Studies (STS) concepts of the 'social life of code', 'data trajectories' and 'data labour' (Bauer, 2008; Mackenzie, 2006; Nadim, 2016) as well as studies of how algorithms come to matter (Amoore and Piotukh, 2015; Introna, 2016; Jaton, 2017; Mackenzie, 2015; Seaver, 2013; Ziewitz, 2016). While much of this work deals with big data, automation, machine learning or large-scale data analytics, we choose a simple risk score with datasets that are relatively small compared to those used in machine learning. In our case, arriba, which in German stands for Absolute and Relative Risk Reduction: Individual Counselling in GPs' Practice, existed in almost identical form as a paper tool before its instantiation as a digital application. This case shows how data are folded into algorithmic tools by ethnographically detailing the processes of translating arriba into a calculative device that is used to facilitate particular kinds of interventions in peoples' lives.

Arriba contains a six-step script for doctor-patient consultations. The formulation of the steps is (in German language) based on the arriba acronym (Table 1). In short, arriba stands for both a risk calculator and a counselling concept.

In addition to the main objective of un-blackboxing risk-predicting algorithms that operate at the interface of population and individual, our article shows how algorithms are embedded in mundane medical and public health practices. We focus on how these mundane practices and tools configure accountability relations, for instance in the negotiation of numbers in the counselling process and in efforts to validate algorithmic tools. In doing so, we demonstrate the salience of broader empirically based notions of algorithms, which are here understood as sets of rules, as step-by-step action regulations, and as systematic instructions for the solutions to standardized problems that are not necessarily solvable by a computer. With respect to medical practice and decision-making, explicit descriptions of how to proceed in a specific situation by following a sequence of

steps, including a branching logic and/or representation as flow chart, are termed 'clinical algorithms' or 'protocols' (Komaroff, 1982). Discussions of these more or less digitized algorithms, their effects, gains and losses, their being tinkered with and their ways of transforming medical work (Berg, 1997a, 1997b), were all linked to issues of health care costs and appropriate care. The ongoing debates that we explore in our study of how epidemiological risk scores 'do accountability' are echoed in current discussions of medical scoring systems, evidence-based medicine and professional accountabilities. Rather than following 'data journeys' (Leonelli, 2014), we follow the mundane everyday circulations and translations of the arriba algorithm (in a broad sense) to grasp specific versions of accountabilities-in-practice.

Unpacking accountability

Accountability procedures in and beyond healthcare settings are commonly associated with performance indicators, quality control measures and audit regimes (Munro and Mouritsen, 1997; Power, 1999). By claiming to provide evidence and enable optimization, accountability practices link ideas of measurability with ideals of good practice. Challenging the assumed split of accountability and care in medical contexts, Jerak-Zuiderent (2015) shows how 'accountable care' in Dutch nursing homes is accomplished and how, in this process, accountability devices are 'a highly singularized form of accountability that always needs to be re-related by specifying from where and for whom and what accounting and caring are done' (p. 430). Such a critical stance towards the taken-for-granted character of accountability, and thorough scrutinizing how it is achieved, captures the 'messiness of accountability in action' (Neyland and Woolgar, 2002: 272) and the social interactions and technologies of rendering things (ac)countable (Law, 1996; Neyland, 2013; Suchman, 1993). These STS approaches to the ways in which accountabilities are created and put to use open up analyses of the various processes and orderings in which risk scores participate. Furthering this line of research, we examine risk scores as devices that produce, maintain, mobilize, and shift practices of accounting and distributions of accountability.

For our purpose of approaching social worlds of accountability in public health, Strathern's (2000: 1) observation of accountability's 'dual credentials in moral reasoning and in the methods and precepts of financial accounting' serve as a starting point. What Strathern discusses for institutional practices in academia holds even more for medical settings, where good practice is not only about professional virtues, benefits and costs but also about well-being and care for people who need support. Since the methods of accounting relate to not only financial practices but also diagnostic/therapeutic methods, we are interested in sites where financial and moral accountabilities intersect and give rise to new versions and combinations of accountability (Reubi, 2018).

Risk calculations in medical consultation tools are such sites where 'the financial and the moral meet' (Strathern, 2000: 1). What is mobilized each time that the risk-predicting algorithm is put to use in a doctor's office? How can we describe the various translational activities that take place around an algorithm in action? Unpacking the circulations of the risk score helps us to better understand how versions of accountability are produced in

public health and how they enact specific politics of health. To this end, we examine risk scores as calculative infrastructures and busy intersections that reconfigure both accountabilities and the relation between collective and individual. Specifically, we focus on how researchers create, adjust and calibrate risk-predicting algorithms for usage, what happens when doctors employ these algorithms in the consultation room, and how risk scores become infrastructures for systematic validation of the score at the population level.

Following the algorithm

Inspired by ethnographies of infrastructure (Star, 1999), we adopt an approach of 'following the algorithm' in order to unpack risk scoring in practice. Taking seriously the Actor-Network Theory enjoinder to follow the actors, including the material and mundane ones, we explore epidemiological risk scores in 'whatever form they might take, whatever they might become and wherever they might go' (Barry, 2013: 426; Latour, 1987). In other words, in following the algorithm in this way, we embark on an 'experiment in contexting' (Asdal and Moser, 2012). Addressing the question of how accounting and accountability takes place in risk scoring, we examine how the score acquires legitimacy and becomes authoritative even while practitioners recognize issues and limitations.

The most popular cardiovascular risk scores in Germany – SCORE Germany, PROCAM, and arriba – are mentioned in national guidelines, commentaries of professional associations or by our medical interlocutors when it comes to the risk-adjusted prevention of cardiovascular disease. This material and comments 'in the field' suggest that cardiologists apparently prefer the ESC-Score, internists PROCAM and family doctors arriba. Arriba is available as free software and allows doctors to create a 10-year-prognosis of a patient's overall risk of a heart attack or stroke. The risk factors considered and processed are: sex, age, smoking, cholesterol levels, blood pressure, manifest arteriosclerosis, anti-hypertension therapy, diabetes and family history of a manifest cardiovascular disease. Besides calculating risk estimates, arriba visually represents the computed probabilities and the impact of risk reducing measures (preventive effects) with the help of an icon array of 100 smileys (Figure 1).

As mentioned, arriba advises its primary users, GPs, to follow a six-step script during the doctor-patient consultation. The German Society for Family Physicians promotes arriba and has recommended it since 2017 in guidelines on cardiovascular prevention in GP risk consultancy (Deutsche Gesellschaft für Allgemeinmedizin (DEGAM), 2017). Medical students learn about it in several universities, and in 2008 the software was awarded the *Berliner Gesundheitspreis* (Berlin Health Prize) as a good means of encouraging cooperative communication between doctors and patients. Further, one of the largest statutory health insurance providers[2] made it part of their GP-centred contracts and forced providers of medical office software to accept an arriba software interface. We chose arriba not just for its catchy name and visual features, but for its significance in general medicine, and more specifically for its usage in medical checkups offered by the statutory insurance providers at two-year-intervals, beginning at the age of 35. Although these voluntary check-ups start with the individual, they can be understood as a

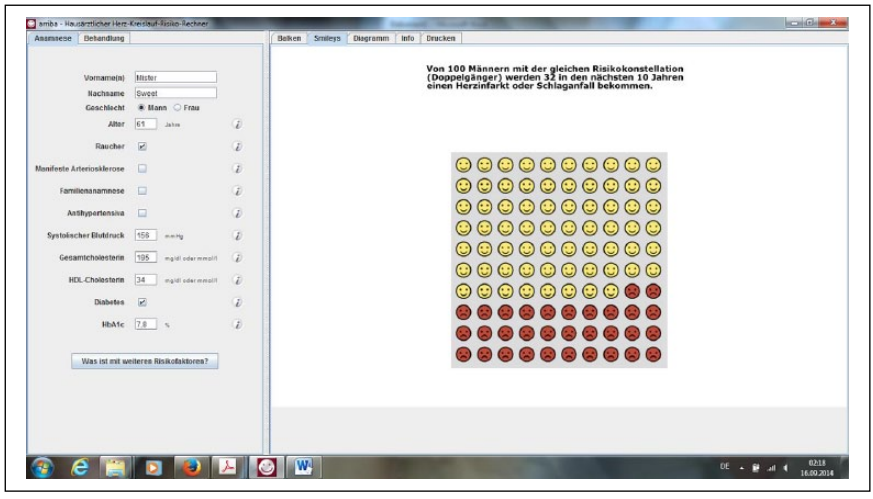

Figure 1. Arriba interface, visual output: smileys (icon array). 'Of 100 men with the same risk constellation 32 will have a heart attack within the next 10 years.' From a screenshot of arriba-software (version 3.03) by GPZK mbH (non-profit society for patient-centered communication ltd.).

population strategy because they target a certain age group to produce assessments of risk (von Lengerke, 2010: 76).

In terms of fieldwork, our approach of 'following the algorithm' can best be described as 'polymorphous engagement' – a term Gusterson (1997) coined when thinking through the methodological challenges anthropologists 'studying up' face with respect to the discipline's defining methodology of participant-observation. 'De-emphasizing participant observation', he proposes a more multi-sited, method-mixing research strategy. This means 'interacting with informants across a number of dispersed sites, not just in local communities, and sometimes in virtual form; and it means collecting data eclectically from a disparate array of sources in many different ways' (Gusterson, 1997: 116). Gusterson's conceptualization resonates with our varying intensive forms of being ethnographically involved when switching back and forth between sites, actors and situations – as well as with Hine's (2007) discussion of multi-sited ethnography and other approaches 'stretching' conventional ethnography in STS (in the context of algorithms, see Seaver, 2017). Importantly, Gusterson's approach also highlights ethnographic knowledge production as something that is characterized by improvisation, pragmatics, priorities and usage of opportunities (c.f. Knorr-Cetina, 1981; Law, 2004). What we follow and how we follow something is as much shaped by research interests and serendipity as by the disciplinary lenses of ethnographers, which in our case is informed by our work at the intersection of STS and epidemiology (Bauer) and STS and cultural anthropology (Amelang), respectively.

Between spring 2014 and winter 2015/16, we followed the algorithm in order to examine its encounters with medical practitioners, patients, health insurers, software developers and epidemiologists. Repeatedly locating the algorithm in the diverse sites

involved in its production and circulation led us to, for example, doctor-patient consultations in primary care in Berlin and the Frankfurt Rhine-Main area, telephone conversations with many different kinds of health-professionals, software interfaces, a crash-course in epidemiology for qualitative social scientists working in public health, and a university class for medical students. We participated in and observed various courses, conferences as well as with three practitioners in their patient consultations. In addition, we conducted in-depth interviews with the three key developers of arriba (two general practitioners and professors of university medicine, one software engineer), two epidemiologists, six users (GPs), three non-users (GPs) and two other proponents of the risk score (non-GPs). We conducted fieldwork and interviews in German and translated the interviews into English ourselves. Moreover, we followed the algorithm through various kinds of documents: scientific journals, medical guidelines, risk charts, health information leaflets, professional journals associated with health insurance companies and doctors' associations, websites, online-discussion forums and extracts of computer code. The three stories we present in the upcoming sections provide insight into our 'polymorphous engagements' with following the algorithm.

First, we open up the black box to re-tell the production of the risk-predicting algorithm and trace how populations become a resource for evidence. Second, we examine the risk score in action in the office of GPs where accountability unfolds at the level of the doctor-patient relationship. As we show below, the individual is enacted through arriba's computation in a specific way. Third, we explore the making of a validation infrastructure by following the practices of retrieving data from counselling to test and improve the score by creating a virtual cohort.

'Adapting Framingham': The arriba score in the making

To introduce the multiple lives and shapes of the arriba risk score, we first consider what went into its development. We unpack how a specific mode of accounting in terms of population health is intrinsic to the algorithm and how this participates in enacting and shaping what we call a 'public health subject' – an individual who is constituted and configured through aggregate data at the population level. When we started our research, arriba had been in existence for twelve years as a 'decision aid' based on an adaption of an established risk formula (Krones et al., 2006; Sadowski et al., 2005). Hence, following the score included locating the risk models and the epidemiological studies that it was based on. Approaching the algorithm's creation involved separate conversations with the three main developers, various epidemiological studies, arriba's earlier paper version, presentations, manuals and reports on the arriba website, and eventually 'seeing' the algorithm at the level of source code. We encountered similar stories of arriba's origin with variations in emphases and perspectives. This points to the processuality in the making of algorithms – especially because they seem to be agents that are powerful yet at the same time, are still in the remaking. We reassemble arriba's development story to provide insight into the negotiations and number-crunching involved in the specific configuration of an algorithm – the crafting and tinkering that is part and parcel of the making of such a tool.

Before arriba was released as software in 2007, its material form consisted of two sheets of paper – the risk calculation sheet and the consultation sheet. The first one (blue for men, pink for women) assigned points to an individual based on the risk factors described above. A 'risk' score was generated from a table. In this way, the complex statistical models, risk functions, and equations that are derived from population studies are translated into a points system, which then facilitates the use of these models in clinical practice. This scoring system, together with arriba's six counselling steps, can be understood as an algorithm in a broad everyday sense. Yet, in the paper version, the actual risk formula – that is, the risk-predicting calculation rule – remains opaque, even though a reference to the source of data is given. Thus, to return to the issue of arriba's making: Who and what else was involved in its production? What are the different translations, pragmatic decisions, and heterogeneous valuations involved when researchers are pulling together evidence for a risk score?

According to the current key developers, 'the arriba team' has comprised at least 30 people based in different cities in Germany in the process of its ongoing making. The second GP and the software engineer joined the team in 2005 to translate the paper-based tool into software. The initial goal was to simply count the points electronically, but it soon grew to make direct use of the multifactorial statistical models obtained from large epidemiological observation studies. To estimate the risk faced by a healthy individual, developers of arriba's first version – the paper version – drew from the *Framingham Heart Study*'s risk score function and variables (Sullivan et al., 2004). This famous longitudinal survey, which was referred to as the 'mother of all cohort studies and classical cardiovascular risk factors' by a GP developer to medical students during a class, started in 1948. Can a tool used in 21st century Germany really be based on an 'old' population study from the US? To solve this issue, arriba's developers balanced the two criteria of 'best validated' (data from the Framingham study) with 'close to context' (smaller datasets for a study population from Germany or Europe).

When we observed its use in consultation rooms, it became apparent to us that the score actualizes past modes of population health counting and accounting that are based on large-scale attempts to study 'lifestyles' as medical risk factors in the second half of the 20th century. However, the more lifestyle variables that epidemiologists try to combine, the more these variables fail to fit into one model, and the more likely the possibility of overestimation (Altman and Royston, 2000; Lenz and Mühlhauser, 2004). As two of our interlocutors pointed out, recently developed biomarkers neither significantly increase the predictive power of the classical cardiovascular risk factors nor improve the quality of existing risk scores (see Tzoulaki et al., 2009; Wang, 2008). Because of this, and because of the pragmatic need to keep things simple, arriba developers considered Framingham's database to be the best despite its shortcomings.

There is, however, a broad consensus that the Framingham risk function overestimates the event rate for European populations. In Germany, for example, researchers tested this with data from the *MONICA-Augsburg Study* (Southern Germany), which was part of an international multi-centre study managed by WHO (Tunstall-Pedoe, 2003), as well as PROCAM (Hense et al., 2003), which stands for Prospective Cardiovascular Münster (North-West Germany) and is the basis for an identically named scoring system (Assmann et al., 2002). Subsequently, the arriba developers adjusted the risk algorithm

with the help of yet another validation study that tested the predictive accuracy of the Framingham risk score: the *British Regional Heart Study*, which in 1978 enrolled 7,735 men from 24 British towns (Brindle et al., 2003, 2005). The arriba developers used the British data and risk estimates to adjust the Framingham risk formula/algorithm with a regional calibration factor, i.e. 'multiply by 0.68'. Then, to add a factor for the event stroke (something GPs involved in the early arriba-testing asked for), they drew on data from the above-mentioned MONICA-Augsburg study that monitored cardiovascular disease in 10,000 inhabitants of Bavaria from 1978-2002 (Holle et al., 2005; Löwel et al., 2005). Finally, the developers added two other factors to capture family history and to take the specific risk of persons with diabetes into account. The latter factor they obtained from a meta-analysis of ten cohort studies with diabetics (Selvin et al., 2004). In short, and in contrast to other epidemiological risk scores based on a singular population study, 'arriba is not a unique risk formula', as one developer emphasized. To adapt the Framingham score, the developers needed to tinker with the data and risk estimates from different studies. Recombining and adjusting these figures, they decided to add four simple coefficients to the original risk-predicting algorithm from Framingham. It is in this ongoing calibration of the risk prediction that versions of accountability are negotiated and become visible and that the 'accountability assemblage' of risk prediction can be opened up for debate (Hoeyer et al., 2019; Hogle, 2019).

In addition to the algorithm's development, the visual presentation of its output was an equally important part of making up the software. One developer criticized the symbols used by other programs because they were too judgmental: traffic light colours to indicate a low or high risk (Score Germany), or a tachometer (PROCAM-score). Such visual design elements run counter to developers' concern for building software that supports discussion with patients about their perception of the calculated risk. While the practice to impart quantitative information with not only words and numbers but also pictures has its own history in graphic design (Tufte, 2009), which graphs best enable the understanding of probability remains an open question for the field of health risk communication (Ancker et al., 2006; Goodyear-Smith et al., 2008; Spiegelhalter, 2008). When building the software, developers refined visualizations of individual risk estimates that had already been an integral part of the paper-based tool. Besides arriba's famous smileys and the invitation to 'picture 100 imaginary doubles who are similar to you in age, sex and risk profile', the paper tool supplemented the score tables with a percentile curve (median, 5th and 95th percentile). The developers had asked the *German National Health Interview and Examination Study* from 1998 to calculate age percentiles in order to allow comparison of individual risks with that of the German population. The software introduced the option to visualize the algorithm's output as a bar chart (Figure 2).

While the invitation to picture 100 imaginary doubles of oneself encourages individuals to take on a population perspective, the bar chart comparison relates individuals to statistical measures vis-à-vis averages. The bar chart view was controversial in the team during software development. One critic professed: '[To this day], the bar charts do not mean anything to me, I have no use for them. Nobody in my local team liked them.' The supporting developer admitted that the bar chart is 'not for purists' and in no way appropriate to demonstrate probabilities. Nevertheless, he assured that it is a 'trick' for people

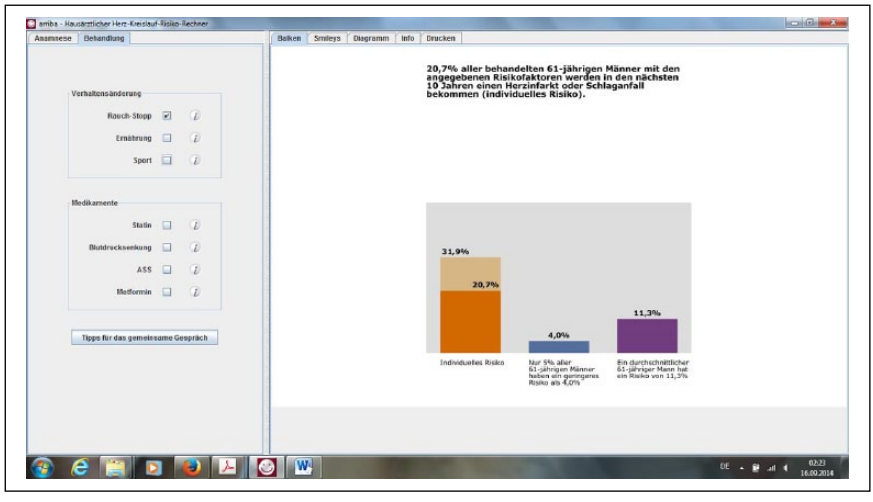

Figure 2. Arriba interface, visual output: bar charts. Risk computation with the arriba-software, including the impact of one preventive measure (stop smoking). The bar chart view allows comparing the individual risk (x% of all patients having the same age, sex and risk constellation) [left bar] with the risk of an average person in Germany of the same age and sex [right bar] and with the healthiest 5% of this group [middle bar]. Screenshot of arriba-software (version 3.03) by GPZK mbH (non-profit society for patient-centered communication).

who are not used to probability thinking, a didactical trick ('Let's pretend as if a probability is a length …') to help present the risk computation in a similar format to the kinds of predictions that routinely circulate more widely in Germany, e.g. in the news. Icons like smileys are a new pictorial language. The icon array emphasizes frequencies, a visual feature developed for a public not used to probabilities. This part of software development is not simply about providing nice graphical features or visual representations of the calculated numbers of evidence-based statistics. Translating probabilities in percentile curves, bar charts or icon arrays draws as much on discussions of appropriate strategies of health risk communication (e.g. Fagerlin et al., 2011; Garcia-Retamero and Cokely, 2013; Kurz-Milcke et al., 2008) as it adds a visual form of persuasiveness to the output in numbers and thus a means to make population-derived risk knowledge count (for the individual). This illustrates how modes of doing accountability are embedded in software. While subject to controversy among developers, once circulating in software packages they become part and parcel of calculative infrastructures that are no longer up for debate.

As a follow-up to the visualized probabilities, arriba allows for demonstrating possibilities of preventive intervention (e.g. first bar of Figure 2). The calculation of the relative reductions in risk that are associated with preventive measures, such as regular physical activity, smoking cessation, intake of cholesterol-lowering or antihypertensive drugs, was a feature of the paper version. Yet, the software not only counts automatically but also foregrounds and speeds up the process of assessing the key issue, namely, how can one influence the risk prognosis by lifestyle changes or by taking a certain drug? It

enables users (doctors, together with their patients) to quickly generate and compute numbers, to visualize them graphically and to 'play around' with modifying this or that risk factor, or multiple risk factors, and 'see' the resulting risk reduction. The effects of changes on a risk factor have been modelled by the developers on the basis of reviews of cohort studies and randomized clinical trials and have been updated for new results, e.g. on diet or specific drugs. Even without these calculations of risk reduction 'the formula we have now is much more complicated than the score charts', notes one developer. He continues, 'as a result, the progression can now be displayed more smoothly [than with the rough jumps inherent to the generalizing categories of score charts]. Epidemiologically speaking this does not matter; it makes sense for the knowledge transfer …'. For the developers, visual design is a matter of communication with patients and the facilitation of multi-factorial risk prognoses. Nevertheless, a perhaps-unintended consequence of this didactic practice is that it promotes a specific mode of population thinking; it endorses the idea that populations contain information about individual futures (cf. Armstrong, 2017). In particular, arriba's visual representations habituate users to this practice of extracting guidance and orientation from the population level.

We have thus far outlined the many considerations and decisions that went into developing both the score and the software, or rather the adaptation of the algorithm and the visual representation of probabilities. Our description of these two approaches shows the tinkering required with the 'evidence-base' in the making of the score. How can we make sense of these adjustment practices and negotiations? Drawing from debates within epidemiology we note that such extrapolations from population to individual are controversial among epidemiologists but common in practice (Rockhill, 2005). The score leads to the fabrication of specific entanglements of individuals and populations in temporal-spatial constellations. It performs the construction and reshuffling of population data from the US, UK and Germany, using them as resources. Epidemiological findings from different localities, times and aggregated individuals are translated, configured and folded into knowledge for use by GPs with their patient populations. Here, the professional accountability of the developers of the risk score is linked to the standards and rules of epidemiological science and the craft of counting and estimating; it is about making and stabilizing evidence. With all the recombining of studies and factors taking place, accountability becomes distributed in different chains of translations and attempts to integrate them into one score.

Yet epidemiological routines also include the cautious validation and critical evaluation of predictive power. For example, one of the epidemiologists involved in the German score charts of the European Society of Cardiology warned us that arriba's specific adaptation of the Framingham Score had not yet been validated. How accountable is the score as a calculative device when it mobilizes populations from the past to account for the projected futures of individuals? For the developers, because they are GPs, the core of the tool is the six-step risk communication algorithm. That is, they care less about precise numbers (in the output of the risk-predicting algorithm) than about pointing their patients in the right direction. They aim for translations that are easy to grasp. They are not so much concerned with what the probabilities are, but how to make sense of them. At the same time though, there is the pressure to prove and show that such a calculator is validated – and these evaluations and debates continue.[3] Encountering the world of the

developers and the very making of the risk score has shown that a range of stabilization processes take place in its later usage, but also already during the development of the model of calculation and the risk predictive algorithm itself. We now turn to the issue of how the arriba score enters general practitioners' offices, where it negotiates clinical encounters and modifies counselling.

'Like a stethoscope': The score as working tool

While doctors directly invoke arriba's script when consulting with patients, arriba's risk-predicting algorithm mostly works discretely in the background. The actual risk calculation is the third of six steps of the script that developers suggest that doctors follow. Using the software transforms the doctor-patient encounter into an arriba consultation, which enrols doctors and patients in risk factor- and population-thinking but also in shared decision-making. Despite arriba's prominence in general medicine, it is not clear whether and how many doctors actually use the software (Müller-Riemenschneider et al., 2010). Following arriba into the offices of GPs, this section outlines what arriba does for doctors and patients, as well as for the performance of accountability in the doctor-patient relationship.

The doctors who we interviewed stated they would consult arriba predominantly for patients between the age of 50 and 65, as well as when they have the impression, due to a specific risk factor, that someone deviates from the average. They find the risk calculator particularly useful for persons where it is not obvious whether they should be prescribed medication. They use it equally often to reassure patients who react with panic to news of increased cholesterol levels and to respond to inquiries made over the Internet. These two patient groups coincide with the two target patient groups that the developers invoke didactically in their teaching materials: 'Mr. Sweet' takes it too easy, potentially missing out opportunities for prevention; 'Ms. Worry' puts too much pressure on herself and is rather too anxious.[4] As one developer explained, these two figures symbolize 'the paradigmatic dilemmas of GP care' that arriba aims to resolve. At the level of public health, they indicate populations who are targeted and who are addressed by preventive check-ups and individual risk-counselling sessions. With respect to the question, 'what is arriba?', two images recurred in the responses of the GPs we interviewed. They compared arriba to 'a tool like a stethoscope' and they marked it as metaphorical 'crutch' that works as an aid or a workaround. Although this might be not surprising in itself, we would like to take up these images to scrutinize what GPs figuratively inoculate and work around with arriba. How is this tool aligned with their everyday practices and the enactment of accountability, and what does the algorithm do in this context?

One doctor, who referred to himself as 'a fan since the very beginning of arriba' became interested in the paper tool as 'a good corrective' to the sole focus on single risk factors. Other physicians liked the software as a 'counting tool' that relieves doctors from doing intuitive or mental calculations, 'because you cannot trust your gut when it comes to the different combination of risk factors'. Along with its 'appealing graphical representations', possibilities for getting risks across and its option to calculate preventive effects, GPs emphasized that arriba is 'more than a mere risk calculator'. As 'a comprehensive means of risk communication', it provides the opportunity to discuss

possible actions in form of potential therapeutic or preventive alternatives. Altogether, they viewed arriba as 'a resource' that backs them up with numerical evidence, including additional information from current studies.

Several GPs mentioned arriba's icon array or bar chart view as a means to translate and explain probabilities. This means clarifying the uncertainties of statistical knowledge for individuals, particularly the fact, for example, that they cannot tell their patients if they will be among the 32 affected, red, non-smiling smileys or among the 68 non-affected, yellow, happy ones (Figure 1). Sometimes a patient solves this difficulty of extrapolating from population to individual risk confidently for himself: 'I don't care about the sad ones, that [pointing to the red smileys on the screen] is not where I see myself.' In our observations of consultations, patients typically followed the calculations with keen interest, though not all do, according to the GPs. The patients we met told us in informal conversations that they felt 'more involved' as patients. They said that they viewed the numerical and visual demonstration as 'a good hint', 'nice gimmick' or 'orientation guide'. While they appreciated arriba's output as 'food for thought' – they usually got a print-out with the numbers and graphics to take home – they often did not consider it as an incitement to action. This might be explained by the different character of the suggested recommendations – one involving drug therapy, the other behavioural changes – or by the practical difference between deciding for preventive intervention and living up to it. During our research, another GP-based project (independent of the arriba team) developed a follow-up booklet that gives advice and support to the realization of lifestyle changes suggested by arriba (see Tinsel et al., 2018**).**

In this regard, one doctor prefers arriba as 'a provocative means' for motivating patients to rethink and change their lifestyle. Indeed, some medical guidelines frame the benefits of risk scores in this respect (e.g. German Cardiac Society, 2007). This GP likes to 'shake up' his patients with the 'declining bars', showing them that lifestyle changes are often more effective than medication when discussing risk reduction measures. Here, the use of the algorithm's output shifts from presenting evidence to patients towards directing patients to evidence. The intended surprise-effect works at least for the interaction in the GP's office, provoking such comments as, 'Wow, that much?! Who would have thought it?' Like several other GPs, this doctor also employed the tool to support his position critical of the quasi-automatic prescription of drugs. Here, GPs adopted an 'anti-Big Pharma interpretation' of arriba, which they related to their care for individual patients. Yet, the decision about whether or not to prescribing drugs based on risk scores is in the GPs' office is not only about the benefits and risks (of drugs) for the patients, but also about the GPs' monthly budgets for prescriptions.

When it comes to medication costs, numerical evidence is required by the Federal Joint Committee (G-BA). The G-BA is the supreme decision-making body for the self-administration of physicians, dentists, psychotherapists, hospitals and health insurance funds in Germany. It issues directives for the benefit catalogue of the statutory health insurance funds and thus specifies which services in medical care are reimbursed. Lipid-lowering agents, for example, are only reimbursed by statutory health insurances when patients have a cardiovascular risk of over 20% within the next ten years. Although the G-BA does not predetermine which risk score to use, algorithms such as arriba become devices to account for economic efficiency in the larger context of cost-benefit calculations that are

used to justify the provision of public health expenditures. In addition to the use of arriba as an accounting device that binds GPs, patients and healthcare insurers in relations of accountability, some of our interlocutors bring it up also in relation to the medical profession more broadly. More than one doctor discussed arriba as a 'political tool' or 'strategy' to 'gain recognition' as not a 'barefoot-doctor'. That is, arriba can lend credibility to GPs by providing them a basis of 'real' evidence. Here, the specific context of the German medical field, where practitioners in clinical medicine and medical specialists look condescendingly down on GPs, is important to recognize. While this points to the range of different actors, including insurance providers, with whom GPs feel themselves to be in relations of accountability, most of our material from GPs concerns the distribution of accountability between doctors and patients.

With respect to the role of numerical risk scores in the doctor-patient relationship, one doctor emphatically called arriba 'an additional authority' that shifts the moralizing undertone from doctors to the algorithms of decision-support systems. In his view, a paternalistic style of medicine is no longer effective or appropriate (see also Ariss, 2009; Collins et al., 2007; Stevenson et al., 2000), and numbers are more persuasive. With this shift, a distributed, algorithmic authority enters the doctor's office. The consequences of the algorithm becoming an external mediator and source of authority need to be reckoned with. How doctors employ arriba and deal with its authority in their daily consultation practices is explained as much by firm professional ethos (partly symbolized in the ways the computer screen was set up facing either both doctor and patient or only the doctor while blocking the patients' view) as by circumstantial negotiations. The above-mentioned doctor, who likes to use arriba's numerical and visual evidence to win patients over to a healthier lifestyle, understood the algorithm to strengthen his medical authority.

This idea did however not go unchallenged. For example, during an arriba consultation, the GP and a 69-year old patient compared the risk-reducing effects of physical training and smoking cessation, both of which the patient is attempting. According to the software's calculation, the two interventions had the same risk-reducing impact. Laughingly commenting on the bar charts the patient said, 'I can tell you what I'll do: I will continue smoking and will exercise even more often'. The doctor, who apparently was not discussing the risk factor of smoking with this patient for the first time, was not amused about this kind of bending numerical authority to accommodate unhealthy habits. This case exemplifies two things that we often saw: First, patients interpreted the algorithm's output in their own way and used the evidence provided for their own calculations. This became more evident when interventions were an issue. Second, despite the intention for arriba to support evidence-based decisions made jointly by doctors and patients, not all GPs bought into this idea of shared decision-making, notwithstanding the fact that developers inscribed it into the software via the six-step consultation script. According to one developer, the software works to 'unburden' doctors and 'makes them happy' no matter what a patient decides about keeping or changing risky habits precisely because this decision is with the aid of the software ideally a decision made together. Apparently, not all GPs share this interpretation when it comes to numbers that 'indicate' particular changes in lifestyles and patients who ignore 'medical evidence'. However, in both cases patients are held accountable for the consequences of their health behaviour

or lifestyle decisions. We observed another authoritative effect of the algorithm in patients' and doctors' account of health and health-related practices: The algorithms' specific mode of summing up cardiovascular health into one figure enacts risk as something contained in and distributed across the patient's body and person; a risk score posits an individual risk that can be speculated on and redistributed into actions that optimize health.

One of our interlocutors, a doctor, who finds arriba interesting but does not use it, criticized the 'idea of a healthy life' in the software as 'idealistic', 'unrealistic' and even 'elitist'. She drew attention to the life circumstances and social situations of patients that impede ideas of simply changing one's lifestyle. In contrast to GPs who explained patients' difficulties as 'old habits dying hard', she referred to the low income that makes patients wear themselves out with several jobs and family obligations, and unable to afford fees for swimming pools or sport clubs. Her bringing together of individual risk behaviour and social determinants of health refers to two well-known conflicting and intermingled approaches to public health. One approach holds individuals accountable by promoting individual lifestyle change (Greco, 1993; Rockhill, 2001), while the other advocates for community public health through health promotion (Aronowitz, 2008; Krieger, 2001). This GP's reservation addresses what might not be cast in stone but rather in computer code: a risk factor epidemiology that privatizes risks and translates everyday life into lifestyle and into an individually modifiable risk factor.

Another GP, who does not use arriba and in general takes a critical stance on risk estimation in preventive contexts, learned about arriba from her patients. One came to her with an arriba printout from an internist, and commented tongue-in-cheek: 'I got it in black and white: I am healthy! But my life will end in death.' Despite this patient's nonchalance, the doctor criticizes the fact that risk scores create worries: 'When the number is out there, it sticks and cannot simply be forgotten'; a healthy person is turned into a risk patient, a pre-patient is enrolled into medicine. To think of one's health through numbers instead of through one's bodily experience is, according to another interlocutor, a lesson people actually learned from medicine: 'we, the doctors, trained them to care about numbers'. Although he thinks 'algorithmic thinking destroys sensing one's body and health as well as trusting intuitive truths', he regularly uses the software to talk about cardiovascular risks and prevention with his patients.

All of the GPs we talked with emphasized that 'family medicine should not limit itself' to evidence-based practice and its specific rationalization of risk. One GP commented, 'Or else the distinctiveness of family medicine would get lost, namely to talk with patients about their life situation.' In their view, risk scores cannot live up to social reality, or the heterogeneity of everyday life, because epidemiology assumes everyone to be the same. As the GP said, from the population perspective of epidemiology, 'It doesn't matter who is sitting there.' That is, the details of patients' lives – where they come from, how healthy they feel, and their own intuitive calculations of risk – do not matter. According to this GP, 'It only matters that it is a 60-year old man. I don't care who you are – this is the message of evidence-based medicine. It would be disastrous if I, as a family doctor, would communicate this to my patients.' Despite such critiques of evidence-based medicine, many GPs classified arriba more positively, namely, as a tool to 'support GP-style thinking'. As generalists, they claim that the 'art' and 'science' of

medicine includes switching between different styles and repertoires of medical knowledge. Epidemiology's knowledge base is only one of them (Berg, 1997a, 1997b; Timmermans and Kolker, 2004: 188). The mode of accountability that is folded into the risk score relates to epidemiology, which is accountable toward a population-level evidence, rather than to the individual. Hence, through the risk score, population-level epidemiological knowledge enters the patient-doctor relationship and practices of counselling individuals, and adds to the versions of accountability enacted in the GPs' office.

As a 'stethoscope', arriba does not listen to the sound of a constricted vein in an individual's body but to the sum of hearts of individuals aggregated into a population study, and thus to patterns in epidemiological data. Despite the intimacy of the consultation room, predictions made by arriba do not pertain to the individual but to an aggregate of people sharing the same risk profile. Treatments begin with assessments of relevant populations which are then made relevant to the individual in the consultation room. Like a stethoscope, a risk score cannot represent reality but needs interpretation. It is only 'tentative', 'but it is the best crutch that I have and that's why I use it', as one doctor put it. For the GPs we met, giving a prognosis, numerical and visual evidence are only one part of the aid this 'crutch' offers in their caring about difference and individuality. For most of them, arriba's evidence-base was not an issue; a few simply assumed that a 'German' score draws on German population data. The tool's six-step script provides a plan that many doctors regard as a way to enact accountable and responsible encounters with patients because it combines the production and translation of probabilities with a way to talk with patients about their perceptions, concerns and potential consequences in their everyday lives. Like a stethoscope, arriba is for our interlocutors just one of many tools among which GPs can and do choose in their counselling practices.

While the check-up, with or without arriba, connects evidence-based prevention and a population strategy with doctors' individual care for patients, the doctor-patient encounters that we observed often invoked cardiovascular health problems as an individual rather than a societal risk. As a result, the individualizing focus meant taking both individuals and their social situations into account and, albeit to differing degrees, counting on individuals being interested in the current imaginary of a 'preventive self '(Niewöhner et al., 2011: 725). The patients and doctors we spoke with know that a general interest in numbers, the acceptance of a risk-factor mode of thinking or the 'well, I could get the bike out of the basement, maybe' uttered in the consultation room differs from truly changing one's lifestyle. Patients living in different social conditions were as stubborn with their routine or as reluctant about giving up fond habits as they were playful with the population numbers made relevant to them. According to GPs, what they as family doctors possess simply is 'time'. In comparison to specialists, they usually can follow-up with their patients over longer periods and thus can 'wait and see' what the future will bring for the specific patient. This attitude can also be understood as a way to deal with the imprecise translation from population data to the individual.

When a risk score enters the GP's office, more than one accountability agenda is negotiated in the doctor-patient-relation. Here, practices of epidemiological counting were about the negotiation of what counts. What is negotiated in this process, as we have shown, are the many translations, detachments, (re-)appropriations and politics of a

score – especially in a setting that is more about health promotion and the prevention of potential illness than the curing of existing disease.

'In-vivo-validation' of the score: Retrieving a virtual cohort

Following arriba further in its circulation, this section moves from its role in patient counselling to its integration and validation in preventive medicine and healthcare provision. In particular we are interested in validation as a site of challenging and stabilizing the score. Validation practices integrate the score into regimes of perpetual evaluation processes, enacting the score as part of good care and good medical science (Pols, 2012; Thompson, 2013). We argue that arriba intersects with negotiations about how to account for and allocate resources in preventive medicine, while it remains 'in the making' and is subject to evaluation and continued reworking and tinkering.

While arriba has been stabilized to a certain extent, its use in general practice faces on-going challenges. To enable circulation of the risk score, especially in an era of evidence-based medicine, it needs to align and combine with existing infrastructures and practices of health care provision, while justifying its knowledge base and proving its usefulness. When arriba enters the world of everyday healthcare provision, it encounters a vast number of contingent regulations, demands of justification and public, semi-public and private institutions. Proponents of arriba are faced with 'validation pressure' to prove the calculations are sound and valid for the population being counselled, especially when the underlying calculations are known to be based on geographically distant Framingham data. To address this validation pressure, researchers draw on arriba-in-implementation to create another layer of data collection and harvest a virtual cohort to test the predictive algorithm with the incidence of disease in the counselled population. The possibility of such data collection, however, depends on the organization and data practices of the health care system. The decentralized and dual healthcare system in Germany grants substantial autonomy, within economic restrictions, to licensed GPs to practice medicine in a variety of ways that does not lend itself to the standardization and evaluation of health care provision.

Since medical practitioners may or may not have encountered arriba and may or may not use it, validation efforts have concentrated on setting up a local cohort within which the score can be validated. We take a closer look at a regional validation effort that has worked on integrating scores into GP office software in Baden-Württemberg (Diener et al., 2013). The 'arriba-pro' study uses statutory insurance infrastructures to create a sample for testing the score. The *Allgemeine Ortskrankenkasse* (AOK) operates by federal states or regions and, in rural areas, often serves the larger parts of the population. Within the decentralized statutory insurers in Germany, AOK is known as an insurance company that allows anyone to become a member but provides fewer benefits. For arriba-pro, AOK integrated the arriba calculator into the organizational software of medical practitioners together with a mechanism of data reporting. This was integrated into a program of GP-centred care, for which these general practitioners registered and gave consent (Freytag et al., 2016; Gerlach and Szecsenyi, 2013a, 2013b; Schneider, 2016). In this model, all the medical services are coordinated by one GP who guides patients

through the system of specialized physicians. The program assigns one GP to see their patients through the entire healthcare system. While it was implemented in 2008 with AOK in Baden-Württemberg, such gatekeeping models were controversial and were not accepted in other regions.

This specific implementation of arriba made a research opportunity possible that was usually not available to health scientists in Germany, which was to automatically follow-up with patients through their treatment data. As identified by epidemiologists, the fact that there are no central health registries in Germany makes the situation for population health research different from those countries with centralized health systems (Ohlmeier et al., 2015; Ronellenfitsch et al., 2004; see also Hoeyer, 2019). In the context of Germany's decentralized healthcare system, datasets with linked patient records cannot be automatically generated. It thus requires considerable labour to retrieve relevant information while dealing with a range of classifications and duplications. One of the initiators of arriba software referred to the opportunity of collecting data as 'the only chance that we will ever have … to look at how an algorithm would look when we deduce it directly from the consulted population'.

This mode of building a virtual cohort was referred to by the researchers as 'in-vivo-validation'; it began with an initial pilot sample of 3500 participants (Diener et al., 2013). Aware of the fact that predictions are bound to uncertainties, researchers stress the unique opportunities in the validation design with 'the convergence of the research and the application setting' (Diener et al., 2013: 1), as well as its significance for secondary prevention.

This mechanism, which effectively uses an in-vivo validation design to upgrade data collected from the cohort of patients consulted with arriba, iteratively extracts data through a study design that is valued highly for epidemiological knowledge production. Hence, this set of data practices functions in a double mode of both holding the patient accountable to leading healthy lifestyles as advised by medical science, and enrolling the patients consulted into a population-level accountability infrastructure that is used for iterative optimizing of the risk score itself and the governance of population health. Moreover, the practice of constructing validation processes is also a means of securing and documenting physicians' own accountabilities to the profession and to medical science. As one developer pointed out, data from 'the past overestimate[s] risks, because treatment gets better – therefore physicians are on the safe side. Better to overtreat than to fail to treat.' That new risks might be produced by overtreatment is considered less problematic than the risks of non-treatment. In our case, rather than avoiding the potential side effects that might arise from treating risk factors, the most prevalent goal of GPs was to work towards 'empowering' patients through health promotion. This adds behaviour change to the treatment options, a general trend that integrates various agendas and interests, including professional ethics, policies and markets, pharmaceuticals and individual decision-making, into the individualized version of health care that began already with the risk factor concept in the Framingham study (Aronowitz, 2011).

These deliberations on accountable practices validate not only the risk score but also algorithmic authority, the project of preventive medicine, and the risk-factor concept as such. The measure of cardiovascular risk within arriba has, albeit perhaps not explicitly, been taken up more broadly as device that mediates and enacts relations of accountability

in its own specific ways. In the metaphors of our GP interlocutors, the dynamic multifactorial model can be tested and played with by seeing which 'weather fronts' are coming and comparing the efficiency of counter-measures.

> It is like meteorology. Works well at times, at some times not, but it is valuable to have it. And if it really gets worse, you see it much better. Whether there is a little rain or not is difficult, but a real thunderstorm front … you will see. This is, I guess, what epidemiology can do, for the individual and at larger scale. (GP developer)

The risk calculation is a tool to sort the relevant from the less relevant and to decide the likelihood of the presence of an issue that requires tackling. This mode of assessment translates risk calculations into interpersonal counselling by way of the six arriba steps.

As a tool of evidence-based prevention, the score appealed to many younger doctors who appreciate evidence-based medicine as a bottom-up movement toward more transparent and less hierarchical decision-making in the clinic (Bhandari et al., 2004; Weisz, 2005). At the same time, efficiency and economic considerations are key concerns – here, new modes of accountability that incorporate both the idea of prevention and justification of expenditures. The control of costs by recommending changes to individual 'lifestyle' patterns, rather than prescribing medication, also works towards those goals of integrated healthcare insurance. The software used in arriba-pro implementation added prevention vouchers to the handout that visualizes the results of the score in an icon array with smileys to visualize the calculated effects of risk reduction. Here, arriba is less a tool for risk communication and shared decision-making. Rather, it is used a means of controlling expenditures, which becomes another way of valuating the score. In an alignment between insurers, associations of medical practitioners and public health researchers, the data retrieved from the 'population of patients consulted with arriba' becomes an empirical knowledge base for evaluation.

Considering the risk score is in action through the double notion of accountability conceptualized by Strathern (2000), we can see both the redistribution of responsibilities and resources and the resistance to these shifts and reordering processes. We find both dimensions operating in cardiovascular prevention counselling using risk scores, enrolling individuals and policies into both the moral and financial dimensions of accountabilities.

The future-oriented 'arriba-pro' validation loop does this by scaling back up to the population level, translating individual lives and the experience of disease into a model-and-test system at the population level. At this level, public health interventions test not the effectiveness of their measures but the assumptions of their own calculations. In this way, arriba gives rise to a long-term apparatus of perpetual validation that, scaled back and up to the population, becomes part of the infrastructuring of public health and the mode of how good healthcare can be known and debated. The risk score is influential in its distributed ways through networks of data circulations. The score is the lynchpin in a model-and-test system for improving population health and the conditions of possibility of knowing and evaluating public health.

Together with other distributed tools of accounting, arriba has made it into the healthcare system because it aligns well with a range of different interests and stakes in

healthcare policies more broadly. The Association of General Practitioners' agenda to increase recognition through the emphasis on the scientific knowledge base of general medicine is important in this respect. Their ambition to provide coordinated care joins the need to control costs, and individualized prevention resonates with a focus on patient-centred care. Insurers assess the efficiency of the counselling costs. Risk-scoring fits itself into other practices of scoring in everyday lives, with disease prevention and public health as widely accepted goals and goods, as something continually to be kept track of and be worked on at the individual and at the population level. Tools such as arriba animate accountabilities in public health; they are sites where the moral and the financial meet in quests for continued evaluation and validation.

Conclusion

This article has followed a risk-predicting algorithm used in preventive medicine from its development at the intersection of epidemiology, general medicine and software engineering, to its usage in general practitioners' offices, to its validation and its infrastructuring of public health in a much broader sense. Applying the STS approach of 'following the actors' to 'following the algorithm' enabled us to analyse the score in action and delineate how it is embedded in health care and public health related accountability practices. Through our multi-sited approach, we experimented with variations of 'following' the algorithm, focusing on the ways in which it is being gleaned from a recombination of epidemiological studies to enter clinical practice and then looped back into the organization of knowledge-making and healthcare infrastructures. Exploring the translations from population to individual and back in the primary prevention of cardiovascular disease, we have examined how versions and distributions of accountability are invoked and practiced along the way.

Empirically engaging with arriba, its remaking, everyday circulations and translations included encountering more than one 'arriba-algorithm': the scoring system (of point values), the six-step-script guiding consultation (a protocol), the risk predicting algorithm (an equation), or the algorithm(s) for the visualization of probabilities. Each of these instantiations of arriba solves different problems and relates varying actors, agendas and issues in different (not always computable) ways. How the arriba developers tied the different aspects together goes beyond translating epidemiological risk knowledge into computer code, and provides insight into the ideas and infrastructures involved in the making of risk scores and related accountability practices. Seaver, who did ethnographic fieldwork with developers of algorithmic music recommendation systems in the US, proposed a consideration of '*algorithmic systems* as intricate, dynamic arrangements of people and code' (Seaver, 2013: 9). Departing from notions of a singular algorithm, he stresses that algorithms are by no means deterministic black boxes; rather they are 'massive, networked ones with hundreds of hands reaching into them, tweaking and tuning, swapping out parts and experimenting with new arrangements' (Seaver, 2013: 10). What is translated or folded into algorithmic systems of risk scores like arriba are not only the different epidemiological studies and thus bodies and temporalities but also ideas of a healthy life, how to achieve it and how to nudge people into it (with a risk figure and visual displays).

The double meaning of accountability we introduced by referring to Strathern (2000) in the beginning – the meeting of financial and moral issues in the term accountability – plays out simultaneously in the transformation of individual patients into 'public health subjects' resulting from risk scoring and risk factor epidemiology. Risk scoring is a combination of epidemiology and accounting, similar to what Reubi (2018) noted for philanthropic funding in global health. As for the recombination and translation of aggregate and individual, there is the epidemiological accounting, the folding of data in order to make a population useful for predictions on future health events. Here the financial plays out more indirectly, e.g. when health insurance covers the costs of medication only above a certain risk figure. Apart from that, there is a moral evaluation invoked with every risk calculation run in a GP's practice as part of a check-up program in an appeal that these probabilities should count – for the individual and the public health subject.

It is only at first glance that risk scores seem to reduce the complexities of everyday life, as they detach, shift, translate, displace and enact politics of measurement. However, there is a range of ways that they get handled. They also prompt refusals of being accountable in the way the software foresees, such as when doctors find the computational tool irrelevant or patients engage in their own sense-making of the algorithm. Interestingly, the GPs who had initiated the development of the scoring tool emphasize shared decision-making over a science-based model. Bringing the evidence of medical studies into the doctor-patient consultation for them is a means to clarify which intervention makes sense for whom but still leaves room for individual decision-making. To our interlocutors, arriba is meaningful less because it strives toward an exact risk-prediction than because it is a tool for better, less paternalistic communication with patients while navigating different medical repertoires. While evidence-based medicine stands for a reconfiguration of medicine's knowledge base in favour of epidemiology and population health that allows other actors to enter the consultation situation (Timmermans and Kolker, 2004), clinical practice involves tinkering with the protocol and mode-switching (Berg, 1997a, 1997b).

Following the algorithm brings into relief some of these redistributions of agency through risk scores. Our ethnographic take on algorithms shows how risk scores, like other devices, are mutable and how they can be embraced or met with resistance. Despite their increasing authority in the fabric of clinical encounters, risk scores are under negotiation and become interwoven with local specificities of the healthcare system. Risk scores emerge through layers of infrastructures and become active in infrastructuring. They shape future data flows and the orderings in which accountabilities in public health are negotiated. The predictions made each time the risk scores are put to use prompt slight but significant shifts in the doctor–patient relationship and professional identity of the physician; the location of medical experience and knowledge is no longer centred in the physician, but mediated through predictive technologies and aggregated population-level data. This more distributed mode of knowing and decision-making has led to a data-intense configuration of interlaced accountabilities that both GPs and their patients must face.

Acknowledgements

Research for this article was conducted as part of the project 'Epidemiological Risk Scores as Knowledge Transfer Devices' at Goethe-University Frankfurt am Main, 2013–2017. We thank Christine Holmberg, Ute Kalender, Alexandra Widmer, Francis Lee, Jeannette Pols, Tone

Druglitrø, Klasien Horstman, Martyn Pickersgill, Klaus Hoeyer as well as the anonymous reviewers and the journal editors for their feedback and helpful suggestions at various stages of this article.

Funding

Research for this article was funded by the German Federal Ministry of Education and Research as part of its program on Ethical, Legal and Social Aspects of Modern Life Sciences, (BMBF, grant 01GP1301A).

Notes

1. West German medicine took up US epidemiological studies from the 1960s on (Madarász, 2010). With the introduction of public health master programs in the 1980s, some researchers also pursued an agenda in the tradition of social medicine, striving to extend medicine toward its social dimensions through a 'population perspective' (Fülgraff, 1996; Mielck and Bloomfield, 1999).
2. The healthcare system in Germany has mandatory insurance provided by statutory non-profit insurers covering 90% of insured population, the other 10% covered by private insurance (mostly self-employed professionals and high-income clients). Despite the mergers between statutory health insurers in the last two decades, there are still more than hundred statutory insurance organizations that remained, of several thousands of such associations from the 1920s. The degree of coverage by the insurers varies, with, for instance, some also opening up to complementary medicine, where insurance companies also play a role in conducting studies of treatment effectiveness (Weidenhammer et al., 2007).
3. The arriba team has shown in a randomized prospective trial study that arriba is evaluated positively by doctors and patients as a decision aid (Krones et al., 2008). For a general discussion of the validation of prognostic tools in cardiovascular risk assessment and if they are suited in the German context, see Lenz and Mühlhauser (2004) and Grammer et al. (2019).
4. For examples, see the tutorial of the arriba-tool (module for cardiovascular prevention) published on the website of the arriba software. Available at: https://arriba-hausarzt.de/uploads/files/arriba_broschuere.pdf (accessed 2 March 2019).

References

Altman DG and Royston P (2000) What do we mean by validating a prognostic model? *Statistics in Medicine* 19: 453–473.

Amoore L and Piotukh V (2015) Life beyond big data: Governing with little analytics. *Economy and Society* 44(3): 341–366.

Ancker JS, Senathirajah Y, Kukafka R, et al. (2006) Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association* 13(6): 608–618.

Ariss SM (2009) Asymmetrical knowledge claims in general practice consultations with frequently attending patients: Limitations and opportunities for patient participation. *Social Science & Medicine* 69(6): 908–919.

Armstrong D (1995) The rise of surveillance medicine. *Sociology of Health & Illness* 17(3): 393–404.

Armstrong D (2014) Chronic illness: A revisionist account. *Sociology of Health & Illness* 36(1): 15–27.

Armstrong D (2017) Clinical prediction and the idea of a population. *Social Studies of Science* 47(2): 288–299.

Aronowitz R (2008) Framing disease: An underappreciated mechanism for the social patterning of health. *Social Science & Medicine* 67(1): 1–9.

Aronowitz R (2009) The converged experience of risk and disease. *The Milbank Quarterly* 87(2): 417–442.

Aronowitz R (2011) The Framingham heart study and the emergence of the risk factor approach to coronary heart disease, 1947–1970. *Revue d'histoire des sciences* 64(2): 263–295.

Asdal K and Moser I (2012) Experiments in context and contexting. *Science, Technology, & Human Values* 37(4): 291–306.

Assmann G, Cullen P and Schulte H (2002) Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation* 105: 310–315.

Barry A (2013) The translation zone: Between actor-network theory and international relations. *Millennium: Journal of International Studies* 41(3): 413–429.

Bauer S (2008) Mining data, gathering variables, and recombining information: The flexible architecture of epidemiological studies. *Studies in History and Philosophy of Biological and Biomedical Sciences* 39(4): 415–426.

Berg M (1997a) Problems and promises of the protocol. *Social Science & Medicine* 44(8): 1081–1088.

Berg M (1997b) *Rationalizing Medical Work: Decision-Support Techniques and Medical Practices*. Cambridge, MA: MIT Press.

Bhandari M, Zlowodzki M and Cole PA (2004) From eminence-based practice to evidence-based practice: A paradigmatic shift. *Minnesota Medicine* 87(4): 51–54.

Brindle P, Emberson J, Lampe F, et al. (2003) Predictive accuracy of the Framingham coronary risk score in British men: Prospective cohort study. *British Medical Journal* 327(7426): 1267–1272.

Brindle P, McConnachie A, Upton MN, et al. (2005) The accuracy of the Framingham risk-score in different socioeconomic groups: A prospective study. *The British Journal of General Practice* 55(520): 838–845.

Callon M and Muniesa F (2005) Peripheral vision: Economic markets as calculative collective devices. *Organization Studies* 26(8): 1229–1250.

Collins S, Britten N, Ruusuvuori J, et al. (2007) Understanding the process of patient participation. In: Collins S, Britten N, Ruusuvuori J and Thompson A (eds) *Patient Participation in Health Care Consultations: Qualitative Perspectives*. Maidenhead: Open University Press, 3–21.

Conroy RM, Pyörälä K, Fitzgerald AP, et al. (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *European Heart Journal* 24(11): 987–1003.

D'Agostino RB, Vasan RS, Pencina MJ, et al. (2008) General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation* 117(6): 743–753.

Deutsche Gesellschaft für Allgemeinmedizin (DEGAM) (2017) *Hausärztliche Risikoberatung zur Kardiovaskulären Prävention*. S3-Leitlinie. Leitlinie Nr. 19, Berlin. https://www.degam.de/files/Inhalte/Leitlinien-Inhalte/Dokumente/DEGAM-S3-Leitlinien/053-024_Risikoberatung%20kardiovaskul.%20Praevention/053-024l_Hausa%CC%88rztliche_Risikoberatung_kardivaskula%CC%88re_Praevention_29-08-2018.pdf (accessed 2 June 2019).

Diener A, Celemín-Heinrich S, Wegscheider K, et al. (2013) In-vivo-validation of a cardiovascular risk prediction tool: The arriba-pro study. *BMC Family Practice* 14(13): 1–7.

Dumit J (2012) *Drugs for Life: How Pharmaceutical Companies Define Our Health*. Durham, NC: Duke University Press.

Fagerlin A, Zikmund-Fischer BJ and Ubel PA (2011) Helping patients decide: Ten steps to better risk communication. *Journal of the National Cancer Institute* 103(19): 1436–1443.

Freytag A, Biermann J, Ochs A, et al. (2016) The Impact of GP-centered healthcare: A case-control study based on insurance claims data. *Deutsches Ärzteblatt* 113(47): 791–798.
Fülgraff G (1996) Why public health research? *Gesundheitswesen* 58(Suppl. 2): 105–109.
Garcia-Retamero R and Cokely ET (2013) Communicating health risks with visual aids. *Current Direction in Psychological Science* 22(5): 392–399.
Gerlach FM and Szecsenyi J (2013a) Hausarztzentrierte Versorgung in Baden-Württemberg: Gelingt die Überwindung der organisierten Verantwortungslosigkeit? *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 107(6): 363–364.
Gerlach FM and Szecsenyi J (2013b) Family doctor-centred care in Baden-Wuerttemberg: Concept and results of a controlled evaluation study. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 107(6): 365–371.
German Cardiac Society (2007) Risikoadjustierte Prävention von Herz- und Kreislauferkrankungen [Risk-adjusted prevention of cardiovascular disease]. Available at: http://leitlinien.dgk.org/files/2007_Leitlinie_Risikoadjustierte_Praevention.pdf (accessed 8 May 2017).
Goodyear-Smith F, Arroll B, Chan L, et al. (2008) Patients prefer pictures to numbers to express cardiovascular benefit from treatment. *Annals of Family Medicine* 6(3): 213–217.
Grammer TB, Dessel A, Gergei I, et al. (2019) Cardiovascular risk algorithms in primary care: Results from the DETECT study. *Scientific Reports* 9(1): 1101. doi:10.1038/s41598-018-37092-7
Greco M (1993) Psychosomatic subjects and the 'duty to be well': Personal agency within medical rationality. *Economy and Society* 22(3): 357–372.
Greene J (2007) *Prescribing by Numbers: Drugs and the Definition of Disease*. Baltimore, MD: Johns Hopkins University Press.
Gusterson H (1997) Studying up revisited. *PoLAR* 20(1): 114–119.
Hense HW, Schulte H, Löwel H, et al. (2003) Framingham risk function overestimates risk of coronary heart disease in men and women from Germany: Results from the MONICA Augsburg and the PROCAM cohorts. *European Heart Journal* 24(10): 937–945.
Hine C (2007) Multi-sited ethnography as a middle range methodology for contemporary STS. *Science, Technology & Human Values* 32(6): 652–671.
Hippisley-Cox J, Coupland C, Vinogradova Y, et al. (2007) Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *BMJ* 335: 136.
Hoeyer K, Bauer S and Pickersgill M (2019) Datafication and accountability in public health Introduction to a special issue. *Social Studies of Science* 49(4): 459–475.
Hoeyer K (2019) Data as promise: Reconfiguring Danish public health through personalized medicine. *Social Studies of Science* 49(4): 531–555.
Hogle LF (2019) Accounting for accountable care: Value-based population health management. *Social Studies of Science* 49(4): 556–582.
Holle R, Happich M, Löwel H, et al. (2005) KORA – A research platform for population based health research. *Gesundheitswesen* 67(S01): 19–25.
Holmberg C, Bischof C and Bauer S (2013) Making predictions: Computing populations. *Science, Technology, & Human Values* 38(3): 398–420.
Introna L (2016) Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology & Human Values* 41(1): 17–49.
Jaton F (2017) We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science* 47(6): 811–840.
Jerak-Zuiderent S (2015) Accountability from somewhere and for someone: Relating with care. *Science as Culture* 24(4): 412–435.
Keil U, Fitzgerald AP, Gohlke H, et al. (2005) Risikoabschätzung tödlicher Herz-Kreislauf-Erkrankungen. Die neuen SCORE-Deutschland Tabellen für die Primärprävention. *Deutsches Ärzteblatt* 102(25): A-1808–A-1812.

Kickbusch I (1986) Health promotion: A global perspective. *Canadian Journal of Public Health* 77(5): 321–326.

Knorr-Cetina K (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Komaroff AL (1982) Algorithms and the 'art' of medicine. *American Journal of Public Health* 72(1): 10–12.

Krieger N (2001) Theories for social epidemiology in the 21 century: An ecosocial perspective. *International Journal of Epidemiology* 30(4): 668–677.

Krones T, Keller H, Sönnichsen A, et al. (2008) Absolute cardiovascular disease risk and shared decision making in primary care: A randomized controlled trial. *Annals of Family Medicine* 6: 218–227.

Krones T, Keller H, Sönnichsen AC, et al. (2006) Partizipative Entscheidungsfindung in der kardiovaskulären Risikoprävention: Ergebnisse der Pilotstudie von ARRIBA-Herz, einer konsultationsbezogenen Entscheidungshilfe für die allgemeinmedizinische Praxis. *Zeitschrift für Medizinische Psychologie* 15: 61–70.

Kurz-Milcke E, Gigerenzer G and Martignon L (2008) Transparency in risk communication: Graphic and Analog Tools. *Annals of the New York Academy of Science* 1128(1): 18–28.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Law J (1996) Organizing accountabilities: Ontology and the mode of accounting. In: Munro R and Mouritsen J (eds) *Accountability: Power, Ethos & Technologies of Managing*. London: International Thomson Business Press, 283–306.

Law J (2004) *After Method: Mess in Social Science Research*. London: Routledge.

Lenz M and Mühlhauser I (2004) Kardiovaskuläre Risikoabschätzung für eine informierte Patientenentscheidung. Wie valide sind die Prognoseinstrumente? *Medizinische Klinik* 61(11): 651–661.

Leonelli S (2014) What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society* 1: 1–11.

Löwel H, Döring A, Schneider A, et al. (2005) The MONICA Augsburg surveys: Basis for prospective cohort studies. *Gesundheitswesen* 67(S01): 13–18.

Mackenzie A (2006) *Cutting Code: Software and Sociality*. New York: Peter Lang.

Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.

Madarász J (2010) Die Prävention chronischer Herz-Kreislauf-Krankheiten: Bundesrepublik Deutschland (BRD), Deutsche Demokratische Republik (DDR) und Großbritannien im Vergleich, 1945-1990. *Prävention und Gesundheitsförderung* 5(3): 313–318.

Marks HM (1997) *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*. Cambridge, MA: Cambridge University Press.

Mielck A and Bloomfield K (1999) Realm and goals of social epidemiology. A contribution to determining current status in relation to German language discussion. *Gesundheitswesen* 61(10): 445–454.

Müller-Riemenschneider F, Holmberg C, Rieckmann N, et al. (2010) Barriers to routine risk-score use for healthy primary care patients. *Archives of Internal Medicine* 170(8): 719–724.

Munro R and Mouritsen J (1997) *Accountability: Power, Ethos & the Technologies of Managing*. Andover, MA: Cengage Learning (EMEA).

Nadim T (2016) Data labours: How the sequence databases GenBank and EMBL-Bank make data. *Science as Culture* 25(4): 496–519.

Neyland D (2013) An ethnography of numbers. In: Caulkins DD and Jordan AT (eds) *A Companion to Organizational Anthropology*. Malden, MA: Wiley-Blackwell, 219–234.

Neyland D and Woolgar S (2002) Accountability in action? The case of a database purchasing decision. *British Journal of Sociology* 53(2): 259–274.

Niewöhner J, Döring M, Kontopodis M, et al. (2011) Cardiovascular disease and obesity prevention in Germany: An investigation into a heterogeneous engineering project. *Science, Technology & Human Values* 36(5): 723–751.

Ohlmeier C, Hoffmann F, Giersiepen K, et al. (2015) Linkage of statutory health insurance data with those of a hospital information system: Feasible, but also 'useful'? *Gesundheitswesen* 77(2): e8–e14.

Pols J (2012) *Care at a Distance. On the Closeness of Technology*. Amsterdam: Amsterdam University Press.

Power M (1999) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Prainsack B (2014) Personhood and solidarity: What kind of personalized medicine do we want? *Personalized Medicine* 17(11): 651–657.

Reubi D (2018) A genealogy of epidemiological reason: Saving lives, social surveys and global population. *BioSocieties* 13(1): 81–102.

Rockhill B (2001) The privatization of risk. *American Journal of Public Health* 91: 365–368.

Rockhill B (2005) Theorizing about causes at the individual level while estimating effects at the population level: Implications for prevention. *Epidemiology* 16(1): 124–129.

Ronellenfitsch U, Kyobutungi C, Becher H, et al. (2004) Large-scale, population-based epidemiological studies with record linkage can be done in Germany. *European Journal of Epidemiology* 19(12): 1073–1074.

Sackett DL, Rosenberg WMC, Gray JAM, et al. (1996) Evidence based medicine: What it is and what it isn't. *British Medical Journal* 312: 71–72.

Sadowski EM, Eimer C, Keller H, et al. (2005) Evaluation komplexer Interventionen. Implementierung von ARRIBA-Herz, einer Beratungsstrategie für die Herz-Kreislaufprävention. *Zeitschrift für Allgemeinmedizin* 81: 429–434.

Schneider A (2016) Wie effektiv sind hausärztliche Versorgungskonzepte? *Deutsches Ärzteblatt* 113(47): 789–790.

Seaver N (2013) Knowing algorithms. *Media in Transition* 8: 1–12.

Seaver N (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2): 1–12.

Selvin E, Marinopoulos S, Berkenblit G, et al. (2004) Meta-analysis: Glycosylated hemoglobin and cardiovascular disease in diabetes mellitus. *Annals of Internal Medicine* 141(6): 421–431.

Shim JK (2014) *Heart-Sick: The Politics of Risk, Inequality, and Heart Disease*. New York: NYU University Press.

Spiegelhalter DJ (2008) Understanding uncertainty. *Annals of Family Medicine* 6(3): 196–197.

Star SL (1999) The ethnography of infrastructure. *American Behavioral Scientist* 43(3): 377–391.

Stevenson F, Barry C, Britten N, et al. (2000) Doctor-patient communication about drugs: The evidence for shared decision making. *Social Science & Medicine* 50(6): 829–840.

Strathern M (2000) New accountabilities: Anthropological studies in audit, ethics and the academy. In: Strathern M (ed.) *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge, 1–18.

Suchman L (1993) Technologies of accountability: On lizards and airplanes. In: Button G (ed.) *Technology in Working Order*.London: Routledge, 113–126.

Sullivan LM, Massaro JM and D'Agostino RB (2004) Tutorial in biostatistics: Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Statistics in Medicine* 23(10): 1631–1660.

Thompson C (2013) *Good Science: The Ethical Choreography of Stem Cell Research*. Cambridge, MA: MIT Press.

Timmermans S and Kolker E (2004) Evidence-based medicine and the reconfiguration of medical knowledge. *Journal of Health and Social Behavior* 45: 177–193.

Tinsel I, Siegel A, Schmoor C, et al. (2018) Encouraging self-management in cardiovascular disease prevention: A randomized controlled study of a structured advice and patient activation intervention in primary care. *Deutsches Ärtzeblatt International* 115(27–28): 469–476.

Tufte E (2009) *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Tunstall-Pedoe H (ed.) (2003) *MONICA. Monograph and Multimedia Sourcebook. World's Largest Study of Heart Disease, Stroke, Risk Factors, and Population Trends 1979–2002*. Geneva: WHO. Available at: http://apps.who.int/iris/bitstream/10665/42597/1/9241562234.pdf (accessed 7 April 2017).

Tutton R (2012) Personalizing medicine: Futures present and past. *Social Science & Medicine* 75(10): 1721–1728.

Tzoulaki I, Liberopoulos G and Ioannidis JPA (2009) Assessment of claims of improved prediction beyond the Framingham risk score. *The Journal of the American Medical Association* 302(21): 2345–2352.

Vartiainen E, Laatikainen T, Peltonen M, et al. (2016) Predicting coronary heart disease and stroke: The FINRISK calculator. *Global Heart* 11(2): 213–216.

Von Lengerke T (2010) Inanspruchnahme gesundheitsbezogener Versorgung. Forschungsstand in Deutschland und Implikationen für Prävention. In: Heintze C (ed.) *Adipositas und Public Health: Rahmenbedingungen, interdisziplinäre Zugänge und Perspektiven für erfolgreiche Präventionsstrategien*. Weinheim: Juventa, 65–81.

Wang TJ (2008) New cardiovascular risk factors exist, but are they clinically useful? *European Journal* 29(4): 441–444.

Weidenhammer W, Streng A, Linde K, et al. (2007) Acupuncture for chronic pain within the research program of 10 German Health Insurance Funds: Basic results from an observational study. *Complementary Therapies in Medicine* 15(4): 238–246.

Weisz G (2005) From clinical counting to evidence-based medicine. In: Jorland G, Opinel A and Weisz G (eds) *Body Counts: Medical Quantification in Historical and Sociological Perspectives*. Montreal: McGill-Queen's University Press, 377–393.

Will C and Moreira T (eds) (2010) *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*. Farnham: Ashgate.

Wilson PW, D'Agostino RB, Levy D, et al. (1998) Predicting of coronary heart disease using risk factor categories. *Circulation* 97(18): 1837–1847.

Ziewitz M (2016) Governing algorithms: Myth, mess, and methods. *Science, Technology & Human Values* 41(1): 3–16.

Author biographies

Katrin Amelang is postdoctoral researcher and lecturer at the University of Bremen and co-organizer of the Bremen NatureCultures Lab. She works at the intersection of cultural anthropology and STS, with a particular interest in body-technology-relations and the quantification of health. In her current research she takes menstrual cycle apps as an example to examine processes of datafication and the intimate encounter of algorithms and bodily naturecultures.

Susanne Bauer is an STS scholar at the University of Oslo. She is professor at the Centre for Technology, Innovation and Culture (TIK) and head of TIK's PhD track at the Faculty of Social Sciences. Trained as environmental epidemiologist and in the history and sociology of science, she has published widely on risk assessment and epidemiology as calculative infrastructure. Her current research interests include algorithmic cultures, regulation, technonatures and hybrid ecologies at post-Cold War nuclear sites and airports.