“Commitment is a prerequisite for effective governance”



Ud tatue velit aliscillam dionsenis In interview: Prof. Dr Jonas Schreyögg, health economist at the University of Hamburg, scientific director of the Hamburg Centre for Health Economics (HCHE) – which was responsible for the study – and deputy chair of the German Health Council

Techniker Krankenkasse has presented its evaluation of the TK HzV-Bund contracts in 13 German federal states, and a battle of interpretations promptly erupted. Health insurance funds and medical associations interpret the findings as evidence that GP programmes offer no benefit to insured persons. The man responsible for the study disagrees. Prof. Dr Jonas Schreyögg, Director of the Hamburg Centre for Health Economics (HCHE), which conducted the study, explains in an interview with “Monitor Versorgungsforschung” why the HzV does not control the model under investigation, but rather does so “only to a limited extent” , why he considers it a transitional phenomenon in its current form, and how a mandatory primary care model could reshape our healthcare system.

>> Professor Schreyögg, the national chairman of the Virchow-Bund, Dr Dirk Heinrich – whose association has been a partner in the Baden-Württemberg HzV model for 17 years – accuses you of presenting“ only half the*truth”* in your study (1), because it deliberately ignores the reality of healthcare provision in Baden-Württemberg. How do you respond to this?

I reject this accusation. The fact that Baden-Württemberg is not included was not a scientific or, indeed, a political decision, but is solely due to the terms of reference for which we applied as evaluators and were awarded the contract. TK commissioned us to evaluate its TK-HzV federal contract, which it has concluded separately in 13 federal states. It is important to note that the TK-HzV federal contract in the 13 federal states constitutes a standardised and therefore comparable intervention, which was the subject of the evaluation tender we were awarded.

“The real crux of the matter is the structure of the HzV, as it is currently designed.”

And Baden-Württemberg is simply not one of them.

Only to a very limited extent. The reason is that, alongside the TK-HzV federal agreements, there are special arrangements – such as an add-on agreement in Mecklenburg-Western Pomerania, and separate agreements in Saxony-Anhalt and, indeed, in Baden-Württemberg. However, the Baden-Württemberg contract is structured differently from the federal contract and – I would like to add – also differs from the HzV contract of the AOK Baden-Württemberg, to which Dr Heinrich refers. This means that, for methodological reasons, it would not have been possible to include the KV regions of Baden-Württemberg,

Mecklenburg-Western Pomerania and Saxony-Anhalt not have been placed in the same pool as the federal contracts at all, as the contracts deviate from the nationwide structure of the TK-HzV federal contracts and thus represent a different intervention. Anyone who were to lump them together anyway would be comparing the incomparable. Put simply: this is simply a case of comparing apples with pears. Which, incidentally, would do neither model – neither the TK nor the AOK Baden-Württemberg – justice. Personally, I find the AOK Baden-Württemberg’s HzV model very interesting – it is, in effect, an innovation lab where new ideas are constantly being trialled. Precisely for this reason, it cannot be compared with the TK federal contract, which is structured differently in terms of healthcare provision.

TK claims as*a ‘clear*finding*’* that participants are*‘not better guided through the care*system*’*, do not have fewer consultations with specialists or hospital stays, and do not result in lower expenditure. On which of your findings are these statements based – and do your data support them?

When we look at the total population studied, the HzV reduces the number of GP visits in the outpatient sector very slightly. At the same time, however, we see an increase in consultations with specialists; and in the inpatient sector, there is no significant steering effect. We also find no significant positive impact on medication, incapacity for work or sickness benefit. On balance, and when considered across the entire population studied, participation actually leads to increased expenditure across all service areas. The expected mechanism – fewer hospital admissions, fewer unco-ordinated visits to specialists – does not take effect in this form. The situation is different, however, for sub-populations, in some of which we do observe a moderating effect, particularly in the group of those who follow the referral pathway via GPs. However, the reason for the lack of a steering effect most likely lies not with the GPs themselves, but in the design of the model, which is why the GP-led care model (HzV) produces only a limited steering effect.

This is precisely where a conceptual objection arises: your control group is standard care – but this is not a baseline condition, as the principles of GP practice have long been standardised under Section 73 of the German Social Code, Book V (SGB V). Could the absence of a difference therefore be evidence not of ineffectiveness, but of a ceiling effect within a system that is already functioning well?

This is a valid and important question – it leads to the key conceptual point of the evaluation. Our benchmark is deliberately standard care without enrolment in the HzV, as we wish to isolate the additional, causal effect of enrolment. If standard care improves of its own accord over the years – for example, because management mechanisms are also becoming established there – then we will capture this, as we are examining the differences in trends over time between the two groups. A pure ceiling effect, however, would mean that there is nothing left to manage. We do not see this: broadly speaking, we observe no decline in contacts with specialists without a referral within the HzV, nor any decline in hospital admissions. The real problem is that the HzV does not actually hold the decisive lever – binding management – in its hands.

Why not?

Because insured people are not obliged to follow the prescribed pathway. GPs cannot exercise control if patients are allowed to go directly to a specialist. We have included a specific chart in the report showing the proportion of those who actually follow the pathway – and this figure is very

limited. The real crux of the matter, therefore, is the structure of the HzV as it is currently designed. Incidentally, this is not specific to TK; the HzV is structured this way throughout Germany: it assumes that insured persons will allow themselves to be managed voluntarily.

However, they also count more contacts with specialists as a failure of referral management. In a well-functioning system of general practitioners , wouldn't a planned, coordinated referral actually be a mark of quality – and wouldn't the increase therefore be a success rather than a shortcoming?

In principle, yes. Referred consultations are desirable; uncoordinated direct access is the problem. That is precisely why we look not only at the number of consultations, but also at whether they result from a referral. And here we see a slight decline in uncoordinated specialist consultations, but an overall increase in specialist consultations. An international comparison is revealing in this regard: in France, where a specialist consultation without a referral entails a significantly higher patient co-payment, the number of specialist consultations is falling. In the Netherlands, a consultation with a specialist without a GP referral is practically ineligible for reimbursement. However, the German HzV contains neither a financial incentive in this regard nor an exclusion of specialist care without a referral.

You emphasise that, as a researcher, you do not claim to generalise the findings of the TK-HzV evaluation to other contracts. What is the situation?

Both points can be addressed. If one examines the TK federal contracts in detail, their elements are relatively similar to those of many other health insurance contracts, simply because they have become increasingly aligned over the years. The TK-HzV model we examined is therefore not a unique contract, but one that many other health insurance funds have also concluded with the same or similar service providers, and is thus readily comparable. The structural finding can therefore be generalised: without binding commitments, control remains weak.

"One could also describe binding commitments as the real game-changer."

Let's stick with the potential transferability of the findings. On average, TK's client base is younger and healthier than that of the statutory health insurance system as a whole. Does your study not, therefore, focus precisely on the area where coordinated GP care is least likely to deliver additional benefits?

The average age of HzV participants is just under 50. Yes, strictly speaking, TK does not have representative population coverage – we know this from many Innovation Fund projects. However, due to the health insurance fund's strong growth, the TK population has moved closer to a representative structure. The AOK effectively represents the other end of the spectrum; ideally, therefore, one would need to combine both populations. That would be the optimal approach, but it does not invalidate our evaluation, because we did not stop at a broad overview. We delve deeper – into individual age groups and among those with chronic conditions – and see there that the management measures, for example, tend to be more effective among older people than among younger people. That is precisely why I speak of a conditional management effect. Incidentally, an improvement in care coupled with lower expenditure only becomes clearly evident among the over-80s.

Here, too, this is likely to be a fundamental flaw in the model: the HzV was not originally conceived as a pure gatekeeping system, but as a structured chronic care programme. The evaluation design you have used assumes that participants and non-participants are comparable after balancing. However, you also acknowledge that attitude and motivation are not captured. People who are more health-conscious are more likely to enrol – a classic selection bias. How valid is your difference-in-differences approach, then?

Neutralising precisely this selection effect is one of the study's key objectives. We therefore combine entropy balancing with a difference-in-differences analysis, specifically the stacked DiD approach. The key factor is the before-and-after comparison with the control group: if someone enrols because they are more motivated, then that was already the case before enrolment. In that case, they were already co-operating better with their doctor before the HzV. This is precisely what becomes apparent in the model, as we compare the difference in differences between the two groups over time – that is, the difference (between the two groups) in the difference (after versus before). If the motivation was already present beforehand, it can be identified via the pre-intervention trend; if it arises only as a result of the HzV, it is part of the intervention. We take both into account. The prerequisite is parallel trends prior to the intervention; this is the main assumption for the DiD model. Where the trends do not run parallel naturally, we enforce this through balancing – and we also check the robustness using the Rambachan and Roth test.

You have effectively constructed a kind of 'RCT light'. How large is the dataset, and what was the main methodological challenge?

With over twelve million policyholders, TK is Germany's largest health insurance fund; over one million TK policyholders are enrolled in the federal HzV contracts, in which 10,000 doctors are participating. For the evaluation, we were able to draw on over half a million new enrolments from 2015 to 2022. And in the control group, depending on the year, we have at least eight million observations from standard care available to us. The main difficulty is sequential enrolment: policyholders join year by year, not on a specific cut-off date. For each cohort, you need suitable control subjects, and you must not use the same ones repeatedly, otherwise a bias arises. This is precisely what the staggered estimator prevents. To do this, you need a huge dataset – which we have. We have also made our selection process very transparent.

How time-consuming was the evaluation?

It was a considerable undertaking: from the stage of data preparation and analysis onwards, several of us worked on it for around a year. This is not the sort of study you can wrap up in a few weeks – not least because the methodology, involving sequential enrolment and parallel trends, is very data-intensive and computationally demanding.

There is even more data available at the German Research Data Centre (FDZ) . Would that have made your study more meaningful?

Theoretically, yes; in practice, we're a long way from that. As the FDZ is currently structured, access is limited to a 60-day period – such an analysis simply couldn't be carried out within that timeframe. There is also a substantive issue: to carry out a proper evaluation via the FDZ, one would have to examine all the contracts of all health insurance funds and check which ones are truly identical. Many of these contracts are public, but I am not, at least, aware of any overview of all HzV contracts in Germany. We have therefore, in

the appendix to the study, at least compared the TK national contract with the AOK Baden-Württemberg contract. At first glance, many elements appear similar; in detail, however, there are crucial differences, for example regarding remuneration and the specialist doctor programme. A comprehensive overview of all HzV contracts is long overdue – it would significantly increase transparency, as there is a great deal of speculation surrounding these contracts. However, for the FDZ, the consent of each health insurance fund would also be required, as the funds must be identified in order to assign them to the relevant HzV contract. Consequently, an analysis via the FDZ is not realistic.

There are 13 TK federal contracts across 14 KV districts, but there are 17 KV associations. Wouldn't the three regions without a federal contract be the perfect control group?

It would be tempting, but it doesn't work. This is because these regions are also subject to an intervention and are therefore unsuitable as an uninfluenced control group. The only uninfluenced factor is standard care – and this is no different in the four regions than in the federal contract states. However, you are also raising one of the most important points: the selection and definition of the control group. There are three strategies here: only non-HzV-insured patients in non-HzV practices, non-HzV-insured patients in HzV practices, or a mixture of both. We have analysed all three. The differences, however, are so marginal that they are barely noticeable; for some indicators, the results even match down to the second decimal place. All three are, incidentally, published in the study and its appendix.

Dr Albrecht Kloepfer of the iX-Media Institute has reinterpreted the current debate in an editorial: the real question is not 'HzV yes or no', but 'Under which contractual conditions does a Section 73b model have a steering effect?'. Baden-Württemberg works only because the binding Section 73b GP contract is complemented by voluntary specialist contracts under Sections 73c and 140a.



Prof. Dr Jonas Schreyögg has held the Chair of Healthcare Management since 2010 and has been Scientific Director of the Hamburg Centre for Health Economics (HCHE) at the University of Hamburg since 2011. After studying business administration and economics at the Technical University of Berlin, he also completed his PhD there in the Department of Public Finance and Health Economics (Prof. Dr Klaus-Dirk Henke), qualifying as Dr rer. oec. Following his PhD, he was a postdoctoral researcher at Stanford.

I certainly agree with the general thrust of the argument. That is precisely why I say that both evaluations are complementary. A key difference between the TK national contract and the AOK contract in Baden-Württemberg lies not in the GP contract model itself, but in the specialist programme and the resulting greater integration of specialist medical care. We even elaborate on this in concrete terms in the discussion: the fact that the Baden-Württemberg evaluation found a reduction in polypharmacy, whereas ours did not, could be due precisely to this closer coordination between GPs and specialists. Our study is not a verdict on the idea of GP-led coordination, but rather a finding regarding a specific implementation of the contract without any binding obligations. It is precisely this structural insight – that what matters is the binding nature of the arrangements and the integration within the sector – that can certainly be generalised.

Your study is being interpreted politically in both directions. Does that bother you?

That's the nature of the business; one has to view it objectively. I endeavour to present the scientific perspective in a nuanced way before the whole thing drifts into the post-truth era. Health insurers and service provider associations have legitimate interests and agendas. But it is the task of science to draw a clear distinction between what is politically desired and what is empirically proven.

If you were to summarise the key findings: what does this mean for a future primary care system?

The most important finding is, in fact, the concept of mandatory participation, which is already mentioned in the current coalition agreement. This significantly increases the likelihood of a steering effect – both in terms of the quality of care and cost-effectiveness. You could also describe this binding nature as the real game-changer. Internationally, the situation is clear: I know of no system in the world where it is said: 'You are now managing patient flows, but for insured persons this is voluntary.' In concrete terms, for a mandatory primary care model, this means that binding commitment must be established – either through mandatory regulations or through financial disincentives for unmanaged visits to specialists – so that access to specialist care is channelled through the GP. The second point is to focus explicitly on people with chronic conditions and the elderly, for example by integrating disease management programmes more closely and,

such as the PRISCUS list. Thirdly, there needs to be greater networking within the outpatient sector – including standardised digital exchange of medical findings and digital referrals, which the GP prioritises according to urgency.

In your study, you often refer to examples from abroad. Which specific elements from which countries would you adopt?

Those from the Netherlands, England, Denmark or France. What these well-functioning systems have in common is their binding nature: in the Netherlands, registering with a GP reduces co-payments and makes it easier to book appointments, whilst access to a specialist is only possible via a referral; in Denmark and Norway, access is similarly linked to a referral. For me, this implies a compulsory registration requirement – which is already laid down in SGB V but is not put into practice. It strengthens the management of care, improves doctor-patient communication and reduces du-

plicate consultations. It is also common in such systems to exempt certain specialist areas, such as gynaecology, paediatrics or ophthalmology, from the referral requirement. As for the HzV itself, I am certain that the current HzV is, in itself, a transitional phenomenon on the way to a comprehensive primary care system. It can also be organised digitally, but please ensure it is embedded within the existing care relationship.

A primary care system requires a governing body that mandatorily integrates general practitioner and specialist care. In Baden-Württemberg, this role is fulfilled by a health insurance fund with a large market share. Who should take on this role nationwide – the KBV or the G-BA? And what will become of the obligation to contract, which currently binds general practitioners to the HzV in structurally weak regions?

I would not place this responsibility centrally within a single government body. My proposal is a mandatory basic primary care model for everyone – for all health insurance funds, all insured persons and all GPs – which, however, is structured on a competitive and contractual basis, just as we see today in the HzV. As has been the case under the HzV to date, the contracting parties for both the basic model and the HzV could be either the Association of General Practitioners or the Association of Statutory Health Insurance Physicians, depending on the region. The ability to differentiate regionally has proved its worth, as regions are, after all, in very different situations. What is important to me is that we should not dismantle the established structures, but rather continue with them. An abrupt removal of retention mechanisms in underserved areas would be counterproductive – the sequence

You deliberately use the term *'primary care'*; the current coalition agreement refers to 'primary care models'. Is this a matter of semantics or a substantive disagreement?

This is a very important substantive point which – I believe – is viewed similarly within the Federal Ministry of Health. They refer to the initiative there as 'primary care' because other professional groups will also be involved. Progressive group practices are already doing this – it's just distributed very unevenly across Germany. A modern primary care model takes the whole team into account, not just the doctor.

The TK is calling for a standardised digital initial assessment as a gatekeeping tool. Critics warn that such triage systems are used less frequently by older people, those with lower levels of education and the chronically ill – in other words, precisely those most in need – and that they exacerbate social inequality. Where do you draw the line between useful telematics and potential risks?

A digital initial assessment makes sense, but we must not view it in isolation from primary care providers – this is precisely what the issue of fairness addresses. We know that 90 per cent of people in Germany use at least one paediatrician, youth health practitioner or GP; if we include those who use gynaecologists in a GP capacity, the figure rises to around 95 per cent. In effect, this means that almost everyone in Germany has a primary care provider they can turn to. If, instead, an anonymous system were set up, I believe – as evaluations from other countries show – there is a risk of increased use of the service; moreover, vulnerable groups in particular would be left out. The 116117 number is justified for emergencies, as is 112, but for non-emergencies I would link the initial assessment to the primary care provider. In many countries where I have lived – such as the USA, England or Singapore – the initial call is first answered by a community nurse at a primary care practice, who asks structured questions and assesses whether care is required and, if so, at what level. This approach must be right: first establish a solid foundation, then incorporate tried-and-tested methods.

How do you measure impact without a randomised trial? DiD with entropy balancing

A difference-in-differences (DiD) analysis does not simply compare those who have enrolled with those who have not, but rather their change over time: how does the intervention group develop before and after enrolment – relative to the change over time in the control group? This allows for the elimination of time-constant differences. The problem is that insured persons enrol in the HzV year after year (phased introduction). Conventional models then distort the estimate because they mistakenly treat those who have already received treatment as controls. The stacked DiD approach avoids this: it determines a separate, unconfounded effect for each enrolment cohort and each year, and uses only individuals not exposed to the intervention for the control group. This is preceded by entropy balancing – a weighting procedure that balances the control group so that its measurable characteristics (age, pre-existing conditions, regional characteristics) correspond exactly to those of the enrolled participants – this is also known as a synthetic control group. It is only through this process that trends prior to enrolment become comparable for certain indicators.

Can this assumption be trusted? The Rambachan & Roth test

Every DiD analysis stands or falls on the assumption of 'parallel trends': without the intervention, both groups would have continued to develop in the same way. This can never be proven, as the hypothetical development of those enrolled without the HzV programme cannot be observed. The only thing that can be tested is the period prior to enrolment – the pre-intervention trend. This is where the method developed by Rambachan and Roth (2023) comes in. Rather than assuming that the trends are perfectly parallel, it asks: to what extent might this assumption be violated before the result is invalidated? From the observed deviations prior to enrolment, a margin of possible distortions thereafter is derived. If the effect persists even under plausible deviations, it is considered robust. The metric $\bar{M}$ indicates the extent to which the effect remains valid following a trend deviation; if this value is too low, the effect is not robust. In the TK-HzV study, the vast majority of indicators pass this test. However, when the groups become smaller – for example, in the case of specialist contacts within the adherent group ($\bar{M}$ = 0.79) – even a single indicator may fail this test; in such cases, no robust effect emerges and no conclusions can be drawn.

Note on interpretation: In this methodology, $\bar{M}$ (Mbar) is the 'breakdown' threshold – the extent of a trend breach relative to the largest observed pre-trend deviation, up to which the effect remains significant. The study classifies values below its required level as 'no robust effect'.

Citation
Schreyögg, J., Stegmaier, P.: 'Binding nature is a prerequisite for effective governance', in 'Monitor Versorgungsforschung' (04/26), pp. 12–17. https://doi.org/10.24945/MVF.04.26.1866-0533.2841

This brings us to the transition risk: in Germany, it is easier to dismantle structures than to build them up. How do you prevent a governance vacuum if the HzV disappears before the successor system is in place – at the expense, of all people, of the elderly and the chronically ill, for whom you are measuring positive effects?

By simply not setting up a parallel system and not allowing a gap to arise. We cannot tell insured people: 'Everyone who is not in the HzV is in a mandatory primary care scheme and must follow the care pathway – whilst in the HzV, everything can carry on as before.' That is not how it is envisaged in the coalition agreement either. The only way forward is to have a basic primary care model that applies to everyone from the outset. A voluntary HzV could then be built on top of this, for example if a health insurance fund wishes to supplement its services with a specialist programme through selective contracts or with elements tailored to specific patient groups, because this enhances the effect and enables it to gain a competitive advantage. The transition would then not be a leap into the unknown, but a phased transition – the existing contracts and relationships would remain the framework.

Where do you currently feel most thoroughly misunderstood?

Most strongly where it is claimed that our study shows the HzV is superfluous. The opposite is the case. I see it as strengthening a future primary care model, as set out in the coalition agreement. The second point is a widespread fear among GPs and their associations that a primary care model would transfer everything into the collective agreement. A primary care model as a basis is much more than a HzV – it is binding and comprises many more elements. And, in my view, both the KVs and the Association of General Practitioners could be contracting parties.

If you had to summarise what you've just said in a single sentence: what is the one condition without which no primary care system in Germany will succeed?

It would make sense to introduce a basic primary care model for everyone. If I had to boil it down to a single word: that one word – 'mandatory' – is absolutely crucial. It is only through making it compulsory that a well-intentioned programme becomes a system capable of improving care whilst simultaneously reducing expenditure on healthcare services. As we operate it today, control is effectively left to the discretion of the insured – this is economically inefficient and unsatisfactory for GPs, because they have a management mandate that they cannot possibly enforce. My proposal: let us establish a binding framework and build competitive elements upon it, drawing on the experience gained from the HzV, including that from Baden-Württemberg, without repeating its design flaws. For me, that would be the real lesson to be learnt from our evaluation.

Professor Schreyögg, thank you for the interview. << The interview was conducted by MVF Editor-in-Chief Peter Stegmaier.

1: https://t1p.de/m0a9b

Treating lung cancer in certified centres

>> Whether someone survives lung cancer depends not only on the diagnosis – but also on where they are treated. This is the conclusion reached by a recent study by a group of authors led by PD Dr Thomas Elter (University Hospital Cologne). The study (1), recently published in the Elsevier journal *Lung Cancer*, is based on claims data from BARMER and analyses the day-to-day care of more than 8,000 patients with non-small cell lung cancer. The results are sobering: Treatment in line with guidelines is by no means a given. And as was also shown in the WiZen study by Prof. Dr Jochen Schmitt, MPH (University of Dresden): the difference between specialised centres and non-certified clinics can literally be a matter of life and death.

The current study examined two patient groups separately. The first comprised approximately 900 patients in the early stages of the disease who underwent surgical treatment. The quality indicator used here was whether appropriate diagnostic imaging – specifically an FDG-PET/CT scan or a combination of scintigraphy and CT or MRI – had been carried out within 42 days prior to the procedure. The result: only just over 61 per cent of patients received this timely imaging. Those who underwent it lived significantly longer, statistically speaking, than those without adequate pre-operative diagnostics. The survival benefit was particularly pronounced among patients who had specifically undergone an FDG-PET/CT, compared with those who had only undergone scintigraphy with CT or MRI.

Certified lung cancer centres performed better in one key respect: their patients were more likely to receive FDG-PET/CT – considered the gold standard – than patients in non-certified centres. However, the certification of the treatment centre had no directly demonstrable independent effect on overall survival – rather, the survival benefit was linked to the quality of the diagnostic procedures themselves, not solely to the location where they were carried out.

Even more pronounced differences emerged in the second patient group: around 1,600 patients with advanced lung cancer who received systemic treatment exclusively. For these patients, molecular biomarker testing is crucial, as only those with the appropriate gene mutations are eligible for targeted therapies – treatments which, according to the study data, can almost halve the survival time compared with non-targeted therapies. The problem: The testing rate in non-specialised healthcare settings was dramatically lower than that of participants in the National Network for Genomic Medicine in Lung Cancer (nNGM). Whilst almost all patients were tested in the nNGM programme, the figure in other healthcare settings was at most around three-quarters.

The authors emphasise that, whilst billing data do not provide complete clinical information and the observed correlations cannot be regarded as proof of causality, they nevertheless offer an unvar-

nished view of the reality of care beyond the large study centres. And this view reveals that: Whilst the quality of lung cancer care in Germany has improved overall, the gap between leading centres and standard care remains alarmingly wide. The study authors' conclusion is correspondingly clear: lung cancer patients should be treated exclusively in certified centres. << **1:** https://t1p.de/hm55e