RESEARCH REPORT

Strengthening Primary Care Delivery through Payment Reform

Options and Experience

Robert A. Berenson
URBAN INSTITUTE

July 2020

Adele Shartz
URBAN INSTITUTE

Roslyn C. Murray
GEORGETOWN UNIVERSITY
ABOUT THE URBAN INSTITUTE
The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people’s lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.
# Contents

Acknowledgments iv  
1. Introduction and Plan for the Paper 1  
2. Primary Care Workforce and Relation to Health System Costs 5  
3. Review of Most Common Payment Methods for Primary Care 11  
4. Recent Payment Improvements in the Medicare Physician Fee Schedule to Improve Value 29  
5. The Unique Challenge of Paying for Telehealth and E-visits 37  
6. Primary Care in US Public Delivery Systems 43  
7. CMMI and PTAC Primary Care Payment Models 47  
8. Selected Primary Care Payment Innovation in Commercial Insurance 56  
9. Primary Care Payment in OECD Countries 59  
10. Review of Core Design Issues in Proposed Fee Schedule/Capitation Hybrid Models 66  
11. Conclusion 76  
Appendix A. Issues in Creating a Logical Typology of Primary Care Payment Models 78  
Appendix B. The U.S. Healthcare Primary Capitation Model 83  
Notes 85  
References 91  
About the Authors 99  
Statement of Independence 100
Acknowledgments

This report was funded by the National Academies of Sciences, Engineering, and Medicine. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute’s funding principles is available at urban.org/fundingprinciples.
1. Introduction and Plan for the Paper

Twenty years ago, Jamie Robinson, health economist at the University of California, Berkeley, began his analysis on how to improve physician payment by stating, “There are many mechanisms for paying physicians; some are good and some are bad. The three worst are fee-for-service (FFS), capitation, and salary” (Robinson 2001, 149). The article primarily called for blending of various payment approaches to balance and moderate the undesirable incentives that would be generated by “undiluted” versions of fee-for-service, capitation, or salary alone. Nearly 20 years after Robinson’s insightful paper, policy makers and clinicians are still searching for improved ways to pay physicians. That quest has been focused especially on primary care clinicians in order to support delivery innovations such as patient centered medical homes (PCMH) and accountable care organizations (ACO), while increasing the primary care practice’s emphasis on population health, social determinants of health, and team-based care delivery. The COVID-19 crisis has also introduced unprecedented challenges for primary care practices along with greater urgency to pay for tele-health and related approaches, substituting non-visit based communication for the dominant mode of reliance on in-person office visits.

In the years after Robinson wrote his analysis, FFS became commonly labeled as “volume-based” payment, even though in European and other Organisation for Economic Co-operation and Development (OECD) countries, it is less pejoratively referred to as “activity-based” payment. Capitation remains a dirty word in the US because of connotations of Health Maintenance Organization (HMO)-based rationing of care—real or imagined—necessitating a new vocabulary, such as referring to primary care capitation as “comprehensive payment,” or “advanced payment.” Whatever the labels, a limited number of payment methods remain that either by themselves or in combination become available payment models as part of payment reform. Our intent is to get beyond the caricatures of payment methods and the need to use euphemistic labels by explaining in detail the advantages and disadvantages of payment methods. This will serve as the basis for exploring how they might be combined to enhance payment in order to improve the value of the care provided.

Despite more theoretical, often creative, proposals and a range of primary care payment demonstrations, we are not much closer to a consensus payment model\(^1\)—whether based on an undiluted method or blended into a more complex payment model—that deserves to be promoted and widely adopted. Indeed, given the growing diversity of primary care delivery models—ranging from
remaining solo and small independent groups, to multispecialty group practices and clinicians employed by hospitals—the cliché “one size does not fit all,” seems particularly apt.

Conventional health policy wisdom in the US, explicitly expressed by the Department of Health and Human Services in 2015, holds that FFS produces little or no value and should be replaced with payment methods on a continuum to “population-based payment,” which has the potential to produce the greatest value. The intermediate methods are often proposed as providing “training wheels” and “on ramps” for providers on the route to population-based payment that includes assumption of substantial financial risk. This paper strongly takes issue with the payment continuum concept reflected in explicit HHS and Centers for Medicare & Medicaid Services (CMS) policy (Health Care Payment Learning and Action Network 2017). In fact, although CMS continues to refer to fee-for-service as devoid of value and emphasizes Center for Medicare and Medicaid Innovation (CMMI) initiatives as the route to enhanced value, since 2013, CMS has improved the Medicare Physician Fee Schedule (MPFS) in many ways, as discussed below in detail. Well-designed fee schedules can also improve value.

Indeed, we believe that every payment method, including fee schedules for primary care clinicians, can be designed to produce more or less value. Improved fee schedules can also provide direction for redesign of clinical services, serving as a more readily accepted transition to alternative payment models (APMs; Liao, Navathe, and Press 2019). Further, virtually all APMs are built on a fee schedule base. Yet, proponents of APMs typically assume that the current design of fee schedules, adopted by Medicare and later used as a model by other public and private payers, is a given, with little or no attention paid to how the fee schedule interacts with more advanced payment approaches and can be enhanced to complement the proposed APMs.

All payment methods have inherent incentives, both good and bad, which cannot be avoided even with an optimal operational design. The inherent perverse effects of payment methods can be muted to some extent through operational design but cannot be eliminated. Accordingly, we strongly concur with Robinson’s argument that payment methods can be blended to “preserve the advantages and attenuate the disadvantages of each” (Robinson 2001, 150).

Plan for the Paper

The purpose of this paper is to review what is known about the different methods for how third-party payers pay primary care health professionals and, in some cases, intermediary organizations to which
health professionals may belong. The paper will not explore how the intermediary organization, whether a small or medium size practice or a large health care organization, compensates the clinicians who are either employed or otherwise affiliated with the organization.

Findings from two recent surveys illustrate the crucial difference between the two different approaches to characterizing how primary care clinicians are paid or compensated. The final Center for Studying Health System Change Tracking Physician Survey published in 2009 found that the most common compensation arrangement for physicians was salary—nearly 70 percent (Boukus, Cassil, and O'Malley 2009). In contrast, an analysis of payment methods used for physicians conducted using the Medical Expenditure Panel Survey (MEPS) from 2010 found that that fee-for-service was the dominant method, constituting 93 percent of physician office visits (Zuvekas and Cohen 2010). Our interest are the payment methods payers—public and private—can use to pay physicians directly or to compensate organizations through which clinicians are employed or affiliated.

The paper initially will review the international literature on the impact of a strong primary care component in health systems on costs and quality. We review the evidence about the percentage of US spending on primary care compared to other developed countries, finding that the comparisons are confounded by different definitions of primary care between the US and the OECD data.

The paper next presents a lengthy review of the common payment methods being used across countries and within the US to pay for primary care, emphasizing the theoretical pros and cons of the different pure methods based on perceived experience and empirical analysis. Consistent with our view that all payment methods can be designed to provide more or less value, we review the substantial improvements that CMS has implemented in the Medicare Physician Fee Schedule, which remains the dominant payment model and is the base payment model against which most current alternative payment models are being promoted and tested. One of the basic flaws in using fee schedules is manifested in the COVID-19-inspired move to virtual care using a range of communication devices. We explore why the approach of “overpaying” for these tele-health services through fee schedules is a good short-term approach to getting money into the hands of financially strapped practices but is not sustainable as a long-term policy. We next briefly summarize the approaches that are used in the most common government programs, including federal qualified health centers, finding a range of approaches and current interest in payment reforms.

The paper then turns toward the past decade’s experience with a range of alternative payment models for primary care, exploring both the CMMI demonstrations most directly affecting payment for primary care and primary care-oriented proposals reviewed by the Physician-Focused Payment Model
Technical Advisory Commission (PTAC), established by the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). A recurrent theme in these demonstrations and proposals is the adoption of approaches akin to primary care capitation, and, in the case of accountable care organizations, population-based risk taking based on global capitation principles. We then review a few of the prominent commercial insurance payment innovations affecting primary care, some done in collaboration with government payers.

The paper continues with an overview of how OECD countries pay physicians, finding substantial variation across countries—some use fee schedules, some capitation, and some hybrids of the two, with a growing use of pay-for-performance. These variations and specific innovative approaches are assessed through focused review of recent innovations that might be relevant for the US in eight OECD countries. The paper concludes by identifying the core design issues in crafting a hybrid payment model relying on both fee schedule and primary care capitation, explaining the importance and choices in designs that would have to be addressed.
2. Primary Care Workforce and Relation to Health System Costs

Primary care physicians are the most frequented specialty, conducting about half of patient office visits (National Center for Health Statistics 2016). However, only about 20 percent of young physicians are choosing a primary care profession, a percentage that has been declining for years (Chen et al. 2013). A recent study found that from 2005 to 2015, primary care jobs grew by about 8 percent, whereas the number of jobs for specialists grew 6 times faster. During this time, the share of the workforce devoted to primary care also declined from 44 percent to 37 percent. Another analysis found that only 32 percent of physicians in the US were primary care clinicians in 2010, but excluded obstetricians and gynecologists from this number (Council on Graduate Medical Education 2010). Accordingly, some predict that by 2030, the shortage of primary care physicians might reach as many as 50,000 in the US.

Non-physician clinicians, including nurse practitioners (NPs) and physician assistants (PAs), have served to supplement patient demand for primary care services. When the researchers included these clinicians in the definition of primary care, the growth in the overall primary care workforce increased to approximately 17 percent, but was still far below the growth in specialists. About 75 percent of NPs deliver primary care, up from only about 50 percent in 2010, and 25 percent of today’s PAs deliver primary care (American Association of Nurse Practitioners 2020). Further, as office visits to primary care physicians declined 18 percent from 2012 to 2016, there was a disproportionate increase in office visits (129 percent) to NPs and PAs in primary care practices (Frost and John Hargraves 2018). However, scope of practice and licensure laws in some states limit the activities of non-physician clinicians and their ability to serve as substitutes for primary care physicians.

A rapid increase in the elderly population in the US began in 2012 when the beginning of the baby boomer generation turned 65. The US Census Bureau projected that the population over age 65 would grow about 5 times faster than that below 65 (Dall et al. 2006). The elderly population uses many more physician services than the non-elderly, increasing demand for services.

Further, a greater proportion of the population are developing chronic conditions that may require more primary care and care coordination than before. As of 2014, 60 percent of US adults had at least one chronic condition, 42 percent had more than one, and 12 percent had five or more. Older adults are also more likely to have multiple chronic conditions. About 81 percent of adults age 65 and older
had multiple chronic conditions, as compared to 50 percent from age 45-64 and 18 percent for adults aged 18-44 (Buttorff, Ruder, and Bauman 2017). In the US, both pediatricians and geriatricians are considered primary care physicians, but are targeted to different segments of the population. Only a little more than 1,800 geriatricians were providing care to traditional Medicare beneficiaries in 2017, a number that has also been declining (Medicare Payment Advisory Commission 2019).

Accordingly, some have concluded that care is increasingly getting more complex for primary care physicians, especially in comparison to specialists. Primary care physicians address more diagnoses and offer more treatment plans per visit than any other specialty. The number and complexity of cases are far greater among primary care physicians as well (Katerndahl, Wood, and Jaén 2015). Further, primary care physicians experience higher percentages of diagnostic cases that are considered “high complexity” than specialists face. The proportion of visits that involve high medication complexity is also higher for physicians in internal or family medicine (Goodson et al. 2020).

International and US Comparisons of Primary Care Spending

The primary care community and other policy makers believe that the disparities between primary and specialty care are importantly due to the longstanding payment inequities perpetuated by the distortions in fee amounts in fee schedules, favoring procedural services and tests over evaluation and management care (Berenson and Goodson 2016; Gold and Park 2016). Some suggest that increasing spending on primary care—through payment reform and rebalancing fee schedule payment distortions—would lead to better quality care and lower costs (Friedberg, Hussey, and Schneider 2010). Starfield and colleagues demonstrated that more comprehensive primary care is associated with better population health and lower costs (Starfield, Shi, and Macinko 2005). Other studies have found that gatekeeping and markets with higher proportions of primary care physicians have lower overall health care spending (Chernew et al. 2009; Gerdtham et al. 1998). In contrast to that research, another peer reviewed study found that countries with stronger primary care infrastructures had higher total health care expenditures, but slower growth per capita. The authors attribute the higher spending to the need for additional resources to support enhanced primary care, as these countries already had higher spending at the beginning of the analysis period (Kringos et al. 2013).

Further, evidence within the US suggests that increased investment in primary care is associated with growth in the primary care workforce and slower growth in overall spending. Rhode Island
introduced regulations in 2010 requiring commercial insurers to increase their primary care spending rate by one percentage point per year. Primary care spending in the state grew from $47 million to $74 million from 2009 to 2014, as a result. Over this time, Rhode Island increased its supply of primary care physicians per capita, while per capita spending grew more slowly than spending in other nearby states (Koller and Khullar 2017).

Commentators often cite data that asserts that European and other OECD countries have much higher proportions of health care spending attributed to primary care than the US and thus, larger investments in the primary care workforce (Koller and Khullar 2017). One prominent primary care proponent claims that OECD countries spend 20 percent of overall health care expenditures on primary care services, in contrast to the US, which has been estimated to spend between 7 percent and 8 percent on primary care (Goroll 2019; Goroll et al. 2007; Koller and Khullar 2017). We estimated that the percentage was about 7 percent (Martin et al. 2020). A more careful reading of the OECD data asserted that primary care services were responsible for 12 percent of aggregate spending in OECD countries. However, even that number is exaggerated because of definitional issues.

Our review of commonly cited OECD data demonstrates the use of different primary care definitions of primary care services and also differs from the common definition used in the US. The US definition is based around the professionals who provide the services (i.e., family practice, internal medicine, pediatrics, general medicine, geriatrics, adolescent medicine, and obstetrics/gynecology) rather than the services themselves (Koller and Khullar 2017).

The OECD noted the challenges in comparing primary care spending in a report from 2016. The report used two definitions of primary care, both of which were based on services provided, not the kind of clinician providing the services, as in the US definition. The narrow definition included outpatient curative and rehabilitative care (while excluding specialist care and dental care), home-based curative and rehabilitative care, ancillary services, and preventive services if provided in an ambulatory setting. The broad definition included outpatient curative and rehabilitative care (including specialist care but excluding dental), home-based curative and rehabilitative care, ancillary services if provided in an ambulatory setting and total preventive services in all settings—including hospitals and LTC facilities (Mueller and Morgan 2019).

Using the narrow definition, primary care accounted for 12 percent of total health care spending across 24 OECD countries (not including the US). When using the broad definition, primary care spending represented approximately 17 percent (Mueller and Morgan 2019). Still, both definitions are
broader than the US definition because the services identified are broader than the services provided by US primary care clinicians.

A different analysis of 22 OECD countries in 2018 used a new definition of primary care that, consistent with the US approach, relies on the clinicians who provide the services, not the services themselves. Primary care is categorized as basic care services—general outpatient (provided by GPs), dental, home-based and preventive care—provided outside of the hospital. This new OECD definition includes dental care, which was not included in the earlier definition.

On average, the 22 countries, which did not include the US, providing the requisite data spent on average about 14 percent of health care expenditures on primary care. However, as this estimate includes spending for dental care, which represents 40 percent of what is identified as primary care—a better comparator with US spending is about 8 percent of total health care expenditures (OECD 2018). The figure is more in line with 7-8 percent, the estimated amount offered by various analysts that the US spends on primary care services (Koller and Khullar 2017). However, this OECD figure might be a modest underestimate because pediatricians, internists and presumably geriatricians are not included in OECD’s definition of primary care clinicians. In most countries, these “specialist” physicians typically do not provide much primary care (OECD 2017).

Across OECD countries, generalists comprise approximately 32 percent of total physicians on average. By the OECD definition, the US has the second lowest share of physicians that are considered generalists (12 percent), while the share of generalists in certain countries that the US commonly compares itself to, such as Canada, France, Australia, the Netherlands and Germany, is substantially higher than the OECD average, closer to 50 percent. As noted above, OECD estimates, in contrast to US estimates presented earlier, do not include specialists that function as generalist, primary care physicians in the US, thus producing a substantial undercount of the US percentage in the OECD data (OECD 2017). Using the American Medical Association Physician Masterfile, we calculated that the US specialty and primary care ratio in 2017 is the same as the OECD average—68 percent specialist, 32 percent primary care (Petterson et al. 2018). In terms of physicians per population, the OECD average was 1.1 GPs per 1,000 and, using the AMA data, we calculated that the US ratio was a little more than 0.8 per 1,000 (OECD 2017).11

Primary Care Spending Percentage

Table 1 includes comparisons of primary care spending by various countries and US states, including the definitions used.
### TABLE 1
Primary Care Spending Rate by Health System

<table>
<thead>
<tr>
<th>Source (Year)</th>
<th>Population Covered</th>
<th>Primary Care Spending Rate (%)</th>
<th>Definitions (Sources)</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Health Service, United Kingdom (2012)</td>
<td>All ages</td>
<td>8.4</td>
<td>Payments to general practitioners (press accounts)</td>
</tr>
<tr>
<td>Ontario (2010)</td>
<td>All ages</td>
<td>8.1</td>
<td>Payments to family physicians only (provincial audit)</td>
</tr>
<tr>
<td>Organization for Economic Cooperation and Development (OECD) average of 24 countries (2013)</td>
<td>All ages</td>
<td>12</td>
<td>“Outpatient curative and rehabilitative care (excluding specialist care and dental care), home-based curative and rehabilitative care, ancillary services, and preventive services if provided in an ambulatory setting” (OECD working paper, 2016)</td>
</tr>
<tr>
<td>Rhode Island (2013)</td>
<td>Commercial</td>
<td>10.6 for largest insurer (vs. 5.8 in 2008)</td>
<td>All payments to family physicians, internists, pediatricians, and affiliated advanced practice providers and for approved “common good” services (health information technology, loan repayment, and practice transformation) (published report, 2014)</td>
</tr>
<tr>
<td>Oregon (2015)</td>
<td>Commercial Medicare Medicaid</td>
<td>10.2 8.9 12.5</td>
<td>All payments for selected services to family physicians, general medicine physicians, pediatricians, obstetrician/gynecologists, psychiatrists, geriatricians, physician assistants, nurse practitioners, and naturopaths and homeopaths (published report, 2017)</td>
</tr>
<tr>
<td>Milbank sample of 9 carriers (2015)</td>
<td>Commercial (PPO) Commercial (HMO)</td>
<td>7.7 8.6</td>
<td>All payments for services provided by physicians in specialties designated by the insurer as primary care (including family practice, internal medicine, pediatrics, general medicine, geriatrics, adolescent medicine, and obstetrician/gynecologist) as well as nurse practitioners and physician assistants (published report, 2017)</td>
</tr>
</tbody>
</table>

Source: Christopher F. Koller and Dhruv Khullar, “Primary Care Spending Rate—A Lever for Encouraging Investment in Primary Care,” *New England Journal of Medicine* 377, no. 18.

Other studies have examined the percentage of total spending for primary care in Medicare. One examined primary care spending in Medicare using two approaches for defining primary care physicians and the services they provide. The narrow definition for primary care physicians included family practice, internal medicine, pediatric medicine, and general practice, while the broad definition included nurse practitioners, physician assistants, geriatric medicine, and gynecology. For primary care services, the narrow definition included Healthcare Common Procedure Coding System codes on professional claims, including evaluation and management visits, preventive visits, care transition or coordination services, in-office preventive services, screening, and counseling, while the broad definition included all professional services billed by PCPs. Using the narrow definition, primary care
spending was approximately 2.12 percent of total Medicare spending. Using the broad definition, it was approximately 4.88 percent (Reid, Damberg, and Friedberg 2019).

The lead author of this paper also has determined primary care spending using a modified Berenson-Eggers Type of Service classification (BETOS 2.0), finding that primary care spending for family physicians, general practitioners, pediatricians, general internists, and geriatricians comprise about 2.5 percent of total Medicare spending (Berenson and Braid-Forbes 2020). The lower percentage of primary care spending as percentage of total in these studies likely reflects the relatively greater use of hospital and post-acute care in the elderly and disabled Medicare beneficiary population than in younger populations.

Primary care physicians have historically been paid relatively more on average—both in absolute wages and in relation to the average wage in the country—in the US than in other OECD countries. An OECD Report on physician remuneration in 2007 found that primary care physicians earned 35 percent more in absolute dollar equivalents in the US than the average income across countries studied. Further, in relation to the average wage in the country, US primary care physicians had the second highest ratio of wages to the average country wage—3.4.—compared to the median of the other eleven OECD countries in the survey of 3.1 (Fujisawa and Lafortune 2008). Although it may be dated, this estimate is the last available we were able to find that compares the differential between primary care income in the US and other OECD countries (OECD 2017).

In summary, while conventional wisdom suggests that OECD countries invest substantially more in primary care relative to the rest of health care spending, our assessments using comparative definitions demonstrate that the US actually spends about the same level as other countries. US primary care physicians earn higher incomes than those in other countries. We cannot fully account for differences in primary care spending across countries due to varying methodologies and definitions used, but the exaggerated notion that primary care comprises either 12 or 20 percent of total spending in OECD countries is inaccurate. The actual percentage appears to be closer to spending in the US but, due to the definitional challenges, this judgment is not definitive.

Of note, our review of papers by the OECD and European Observatory on Health Systems and Policies suggests that other developed countries are similarly concerned about a lack of commitment to the primary care workforce, as manifested by compensation disparities and perceived shortages of primary care.
3. Review of Most Common Payment Methods for Primary Care

Appendix A discusses considerations in organizing and describing payment methods, applicable to the common approaches used to pay primary care in OECD countries. It emphasizes that there is no right way to classify payment methods, as a lot depends on the purpose for which the classification is being used. The discussion in Appendix A also makes clear that the labels used to describe payment methods oversimplify the design complexity that any payment method involves. A prime example are fee schedules, which we use in preference to the term fee-for-service, partly because fee schedules are able to pay for bundled episodes, such as 90 day global periods after surgery and even capitation, such as the monthly payment for renal physicians managing dialysis and, more recently, the Medicare monthly chronic care management fee.

For purposes of this paper, we are using the following classification of common payment methods for primary care clinicians. As discussed earlier, we do not review salary here because this method is not available to payers making payments to practices or other organizations, which in turn compensate their constituent clinicians.

Fixed Payments

- Salary for a health professional
- Lump sum payment to the practice or larger organization

Activity-Based Payments

- Fee schedule for primary care

Population-Based Payments

- Primary care capitation (also labeled comprehensive payment)
- Condition-based episode payment

Incremental Payments

- Shared savings (and shared risk)
- Pay-for-performance
We review payment methods by providing a brief background on the method and the method’s strengths and weaknesses. We note that some strengths and weaknesses of payment methods are inherent, whereas others largely depend on the operational design of how the method is implemented. Design can mitigate to some extent the negative effects of payment methods. For example, fee schedules have an inherent incentive to generate activity, thereby increasing volume and, often, costs. However, the selected design to base relative fees on flawed estimates of relative resource costs contributes to overpayment for some services, increasing the level of overprovision of cost-generating procedures and tests, and underpayment for others, hence resulting in under-provision of potentially cost-reducing team-based care and collaborating and coordinating on evaluation and management services.

We also will comment explicitly on a few payment objectives that have attained recent prominence as central objectives in payment reform and are consistent with the potential role of primary care in producing greater value. In accordance with the admonition to “make a virtue of necessity,” the current COVID-19 pandemic has generated widespread interest in greatly expanding payment for various forms of non-visit-based communication between patients and primary care clinicians and practices. A primary objective of payment reform is to produce a reduction in the rate of increase in spending – the denominator of the value equation. Another current objective receiving attention through payment reform is enhanced team-based care within practices and greater collaboration with specialists and other components of health care delivery. Collaboration with a range of organizations providing important social services is also a recent expectation of primary care practices as is greater attention to social determinants of health. These increased expectations for primary care would require increased resources through payment enhancements. We evaluate the potential effects of these payment objectives in the discussion of specific payment methods.

Finally, in this section of the paper, we do not attempt to identify the various design provisions that could be implemented to address disadvantages. Many have been adopted and/or can be envisioned. We present pure, or what Robinson called undiluted, methods to demonstrate the problems that need to be and likely can be addressed, especially in hybrid approaches that blend different payment methods. We do not review the methods in the order of the classification presented earlier, but rather in order of their current importance as likely candidates to be included in new payment models.
Fee Schedules for Primary Care

A fee schedule is a list of the maximum rate a payer will allow for services, with the definition of services based on code sets such as CPT (Current Procedural Terminology) in the United States and ICD-10 PCS (International Classification of Diseases, tenth revision, Procedure Coding System) in some other countries. Typically, the payment is the lower of the provider’s actual charge or the fee schedule allowance. Most payers determine fee schedules first by establishing relative weights (also referred to as relative value units) for the list of service codes and then by using a dollar conversion factor to establish the fee schedule.

Prior to use of prospectively determined fees in fee schedules, public and private payers in the US used variations of the usual, customary, reasonable UCR methods, in which payment was a function of the clinicians submitted charge and the fees for other clinicians in the defined community. (Medicare used a comparable concept to UCR, called "reasonable, prevailing, and reasonable" - CPR). The UCR approach was criticized as inherently inflationary, incentivizing physicians to continually increase their charges. Moreover, UCR perpetuated distortions in charges by providing better insurance coverage for tests and procedures than for evaluation and management services such as office visits. Eventually, payers came to view predetermined payment maximums as the preferred approach.

Medicare initially based fee maximums on physician charges. The MPFS, introduced in 1992, is based on estimates of covered services’ relative practice expenses, the value of physicians' work as measured by time and service intensity, and professional liability costs. These resource costs are adjusted for differences in input prices for goods and services in different markets, then the total is multiplied by a standard dollar amount—the conversion factor—to arrive at the payment allowance for any practice. Most US payers base their own fee schedules on Medicare's, although they generally use higher conversion factors than Medicare's; commercial insurers sometimes modify fees for particular practices based on price negotiations with individual practices that have bargaining power. The Medicare fee schedule has also been adopted in Medicaid, but usually with lower conversion factors, while mostly preserving the relative values across services.

Typically, fee schedules pay retrospectively for one-time services—a procedure, a test, an office visit—using preset fee levels. However, as discussed in Appendix A, even though labeled as fee-for-service, fee schedule codes can be forms of capitation (e.g., payment for a month of dialysis-related professional services) or episodes (e.g., payment for a 90-day “global” period of post-surgery routine care, a month of complex chronic care management). However, in a fee schedule, these forms of bundled payments identified with a code definition can only be used for services provided to a specific
patient, in contrast to true population-based payment in which the payment supports care to a defined population of patients such that the recipient of the payment determines how to allocate the payment across individual patients.

**Strengths**

- Fee schedules reward activity and industriousness, while facilitating patients' access to care because providers get paid more for doing more.
- Fee schedules are well established, with well-studied impacts and consistent with how consumers make payments in retail markets.
- Payment through fee schedules do not require solo and small clinician practices to engage in more complex payment methods requiring larger infrastructure and additional administrative support.
- Theoretically, the approach can encourage desired behavior by paying more to encourage or less to discourage provision of particular services.
- A fee schedule implicitly adjusts for the different case mixes, by paying comparatively more for sicker patients that need more services; it serves as a crude, but useful, risk adjuster.
- The approach provides payers with data about patient care, which can then be analyzed for accountability, such as establishing performance measures. Although payers can require practices to provide encounter data, claims for payment are more reliable.
- Fee schedules can accommodate elements from other payment reform approaches that are similar to capitation or episode-based payments. Adopting forms of bundled payment within fee schedules may preempt the need to replace fee schedules.

**Weaknesses**

- Fee schedules encourage overprovision of services, because clinicians often determine the need for services and can induce patient demand.
- The method ignores whether the service was appropriate or performed well.
- Fee schedules contribute to care fragmentation because payment for individual services provide no inherent incentive for providers to coordinate internally or collaborate externally with other health professionals.
Fee schedule payments generate a large number of billable transactions, which in turn generate high administrative costs for health professionals and payers.

The substantial number of professional activities that are not codified and covered for payment may be marginalized, as discussed in more detail in Appendix A.

It is particularly challenging to pay for high frequency, brief activities, including phone calls, emails, and other forms of non-visit communication because the transaction costs of billing may exceed the dollar value of the communication and, without cost-sharing, may generate increased costs.

Coding complexity, with US payers relying on more than 8,000 codes, makes fee schedules susceptible to “gaming” or outright fraud, especially for office visits, as reflected in twenty-five years of intrusion by the documentation guidelines. Documentation for non-visit communications would be particularly challenging, offering a ready opportunity for gaming.

Fee schedules are prone to payment distortions over time, no matter the basis for differentiating fee levels. In fee schedules based on relative resource costs, payers need to maintain fees that reflect technological changes and work process improvements that alter relative resource costs. Evidence demonstrates major distortions in current fee schedule relative values in the MPFS, notably, overpaying for procedures and tests and underpaying for evaluation and management services in relation to underlying resource costs.

Financial well-being for practices can be compromised when external events, such as COVID-19, produce a downturn in patient demand for services.

**Condition-Based Episodes**

With payment for episodes, a prospective payment is made for all care a patient receives over the course of a defined clinical episode or period of management, instead of for discrete services (as with a fee schedule) or for all care a patient receives (as in capitation). An episode of care has two dimensions: a clinical dimension, which can represent either the set of services or the clinical condition(s) that comprise the episode and a time dimension that reflects the beginning and the end of the episode. In essence, the approach is designed to transfer financial responsibility for the technical (or performance) risk, i.e., risk related to care production, that is under the included providers’ control, but not the probability (or insurance) risk that relates to the burden of illness and injury in any large patient population. When the episode is bundled across different providers, including clinicians and
facilities, the parties have common financial incentives to control the cost of the bundle, thereby incentivizing greater collaboration.¹³

Most work on episode-based payment has focused on procedures, such as hip and knee replacements, and would not apply to primary care physicians. However, there have been pilots of condition-based episodes that could be a method of payment to primary care clinicians. In this approach, an episode for a reasonably well-defined chronic condition commonly would be the focus of the payment model; for example, the episode payment could include services for patients with ischemic heart disease or diabetes mellitus for a period of time, perhaps as long as a year. Compared to procedure-based episodes, bundled episodes for many chronic conditions could affect much more health care spending and could create much stronger incentives for care coordination across a team of health professionals. Procedure costs could be included within a condition-based episode (e.g., estimates of frequency of stent placements and CABGs within an ischemic heart disease episode), thereby reducing incentives to perform inappropriate procedures, similar to the incentives inherent in capitation. However, in contrast to capitation, because most patients with one serious chronic condition often have other chronic conditions, as well as social and psychological factors that affect care, paying an episode for each one separately would be both challenging and non-holistic.

There are two fundamentally different approaches to condition-based episode payment. In a modestly successful approach tested in the Netherlands, the payment for the condition, in this case diabetes, provides a single episode payment for the evidence-based preventive services that should be provided to a patient for a year. The approach does not include payment for the range of health services a patient with diabetes might accrue over the year. The condition-based demonstrations in Medicare, however, do expect the recipient of the episode payment to be accountable for total cost of care associated with the diabetes diagnosis or even all spending for the patient during the episode period, thereby both broadening the potential impact of the approach but introducing a number of operational challenges.

**Strengths**

- Condition-based episodes internalize incentives for efficiency to the recipient of the payment.
- Evidence-based practice guidelines typically are condition-based and so may be more manageable and acceptable to physicians than capitation for all services.
- Condition-based episodes for well-defined clinical conditions lend themselves to targeted performance measurement.
In contrast to procedure-based episodes, this approach counters the typical fee schedule bias to perform often unnecessary procedures, which if performed would constitute a substantial portion of spending for the condition.

Episodes can be viewed as providing an “on-ramp” toward assumption of risk, moving toward capitation, in this case for well-defined clinical conditions supported by evidence-based guidelines.

Weaknesses

Many conditions—even common ones—are not well defined, offering providers an incentive to “find” conditions in order to receive a prolonged payment for a condition-specific episode, thereby creating the “woodwork effect” (eligible patients coming out of the woodwork to generate a full episode payment); currently condition-based episodes are triggered by an accepted ICD-10 diagnosis code on a claim form, rather than on independent clinical criteria.

In fact, even likely chronic condition candidates for condition payment, such as COPD and CHF are commonly, albeit unintentionally, misdiagnosed. It would be perverse to pay a full episode payment for a condition the patient does not have. Current models of condition-based payment do not require confirmation of the accuracy of the diagnosis.

Most common chronic conditions have a broad range of severity ideally necessitating the need for case-mix adjustment to determine payment levels. Practical alternatives, e.g., excluding patients who have had fewer than a certain number of visits in a prior period, are subject to gaming as long as the recipient of the payment can control the utilization criteria that determines who is included and excluded.

Many patients, especially in Medicare, have multiple chronic conditions. Paying on a condition basis, perhaps to different groups of providers, would not be holistic, possibly counteracting the goal of better care coordination of interacting clinical and non-clinical issues that determine patient well-being.

To address the potential for patients with coexisting conditions to receive non-holistic, separate episodes of care, payers could combine conditions that often “travel together” (e.g., hypertension, congestive heart failure, diabetes, chronic renal failure) into a single payment that providers would be responsible for. However, this combining of conditions approaches the complexity of capitation and would undermine the simplicity condition-specific episode payment is meant to achieve.
In managing a chronic condition, the cost of a procedure typically dwarfs the cost of medical management absent the procedure. A single condition-specific payment, then, would perhaps create a powerful incentive for its primary recipient to not refer the patient for necessary procedures. Conversely, establishing two different payment levels—one with procedure, one without, comparable to how DRGs are constructed, undermines the potential benefit of the condition payment to reduce unnecessary procedures.

Since the condition-based episode payment typically is triggered by a claim for care for the condition, the approach ignores the benefits of “upstream” care to prevent disease progression to the point that it qualifies for a condition-based payment, for example, care that forestalls the onset of end-stage renal disease.

**Primary Care Capitation**

Capitation is a prospective unit of payment per patient, per month or year, in which a payer makes a fixed payment for a defined set of services, regardless of the quantity of services actually provided. This payment approach can be used for an individual health professional, for a group of health professionals for their collective professional services (“professional capitation”), or for provider organizations to assume risk for most health services (“global capitation”). Primary care capitation is a distinct approach with its own characteristics and is not simply a narrower form of the broader capitation approaches. Primary care capitation was a commonly used payment method in the 1980s and 1990s and still persists, albeit less commonly, today, associated with the rise and decline of health maintenance organizations (HMOs). Primary care capitation required insured individuals to select one primary care physician (known variously as a “gatekeeper” or “primary care case manager”) both to provide routine care and to approve referrals for nonurgent or emergent health services.

Actuarial problems complicated the earlier primary care capitation models in that they estimated per member per month (PMPM)\(^{14}\) rates based on an actuarial analysis of fee-for-service claims despite the fact that an HMO expects gatekeeper clinicians to expend additional effort that was not recognized under fee schedules and was therefore not included in their payments. In essence, a monthly payment for all needed care activities should include the 25+ percent of activities that are not paid under fee schedules, as discussed in Appendix A (Goroll et al. 2007). In addition, previous HMO primary care capitation models typically adjusted payments for a patient’s age and gender but not for health status, leading to frequent mismatch between patients’ needs and capitation payments.
Many states in the 1980s and 1990s, then concerned about the theoretical incentive to deny needed services under primary care capitation, considered the method to constitute risk bearing. Most of these states thus restrict primary care capitation for use within an HMO structure, but not in other products, including the more common and growing preferred provider organization (PPO) insurance model, because of their belief that the state had greater regulatory oversight over HMOs than other, looser insurance products (Kongstvedt 2013).

A major variation in how primary care capitation is operationalized is whether the clinician bears risk for downstream spending by other health providers or is at risk only for their own primary care services. The major model adopted by HMOs in the '80s and '90s had risk pools in which the recipient clinician stood to lose funding if their total cost of care for all services exceeded actuarial projections (De Lew, Greenberg, and Kinchen 1992). The approach was not true actuarial risk, in that the practice was not obligated to pay for actual spending generated by other components of the health system, but rather served as a powerful form of pay-for-performance in which their payments could receive a total or partial return of payment “withholds” based on analysis of their risk pool at the end of a reporting period. Of note, common withholds were on the order of 20 percent of the capitation payment level, far more than current typical P4P bonuses and penalties.

**Strengths**

- Primary care capitation places “performance risk” on clinicians, providing them financial incentives to limit provision of unneeded services.

- This method provides a platform for primary care clinicians to take responsibility for reducing overall spending for their patients and/or alternatively limit unnecessary service use generated by specialists on referral, care in emergency departments, etc.

- This payment approach internalizes to the primary care practice decisions over the allocation of activity and costs, permitting more flexibility in individualizing care to meet patients’ needs.

- Assuming the per-person-per-month payment is sufficient to cover activities not traditionally recognized for payment under fee schedules, the practice can organize to provide more team-based care, consistent with the medical home delivery model.

- Payers can support newer forms of communication and care delivery that substitute for the traditional office visit, such as telehealth, as well as phone calls and secure email, much more easily and prudently through capitation than under fee schedules.
- The approach gives payers predictable and capped costs, while providing the recipient clinician a predictable cash flow, even during periods, when demand for in-person services is substantially reduced, such as during the COVID-19 epidemic.

- Primary care capitation is administratively straightforward, although design approaches to address its weaknesses can add substantial complexity. At its simplest, providers receive payment every month for a roster of patients, obviating the need for billing and paying for each service provided, as under fee schedule payment.

Weaknesses

- Primary care capitation has traditionally been adopted in HMOs, including Medicare Advantage HMOs, which can restrict patients' choice via the primary care physician gatekeeper role.

- Where clinicians are at risk for downstream spending performance, absent some form of lock-in that in some way limits patient's freedom of choice to see any clinician, primary care capitation provides accountability for spending without the authority practices need to actually limit unnecessary spending.

- In the absence of potential penalties, such as with a risk pool approach, primary care capitation creates an incentive for primary care physicians to liberally refer their patients to other physicians for services that otherwise would be within the scope of the primary care capitation payment, thereby actually increasing total costs and loss of continuity of care.

- Primary care capitation may lead to stinting on care—particularly care that can be avoided without compromising the patient's well-being in the short term (e.g., disease screening and prevention services).

- In the absence of strong risk adjustment for health status, primary care physicians can "cream-skim"—that is, shun sicker, costlier patients that would take up more time and resources in favor of healthier ones for whom payment would be the same.

- In a pure primary care capitation model, payers cannot use a fee schedule or P4P to promote activities or services they want to encourage.

- While administratively simple, reliance on a single monthly capitation payment provides little transparency to reveal clinicians' activities, making performance measurement and assessment, and risk adjustment for health status virtually impossible.
- Payment per capita provides clinicians a financial incentive to take on too many patients, exacerbating concerns about stinting and over-referral.

- Conversely, primary care capitation assumes statistical averaging of patients with different health care needs, so a substantial, minimum number of patients is needed for it to work from a statistical accuracy standpoint. Thus, payers will need to maintain a fee schedule for primary care payments to some clinicians (as well as for specialists) in parallel with primary care capitation, adding administrative complexity and reducing system savings.

- Direct payments to primary care providers represent a small percentage of health care spending, as discussed earlier. Unless the design includes a strong incentive system for rewarding or penalizing total health care spending or targets of avoidable service use attributable to the physicians’ patients, this approach might not be worth the implementation effort.

- HMOs are subject to state regulatory oversight and therefore can apply primary care capitation. Self-funded employer plans are not subject to state regulatory oversight so they could use primary care capitation, but as a practical matter they typically contract with available state-regulated HMOs (capitation permitted) and PPOs (not permitted) and go along with the payment method the plan administrator uses.

Pay-for-Performance

A pay-for-performance (P4P) model consists of financial incentives or penalties based on a provider's ability to meet performance expectations based usually on publicly available, predetermined measures. P4P provides additional bonuses and penalties in addition to the base payment providers receive. P4P mostly has been associated with providers’ performance in meeting quality of care standards or improving the quality provided to the patients for whom they are deemed responsible. P4P models measure performance using clinical process and outcome measures and surveys on patients' experiences with care. The Merit-based Incentive Payment System (MIPS), created by the Medicare Access and CHIP Reauthorization Act of 2015, a prominent example of P4P, also includes measures of “meaningful use” of electronic health records and resource use.

Typically, performance bonuses or penalties have represented a few percentage points of the base payment providers or health professionals receive. However, because the economics of different providers vary substantially, a 1 or 2 percentage point bonus or penalty has much different impact. For
example, hospital margins are usually in the single digits, whereas physician net income for a small physician practice, representing the physicians’ take-home income, are typically in the range of 40-50 percent of revenues. The same P4P percentage of incremental payment, thus, might impact hospital behavior much more than physician behavior, although physicians might have a greater interest in achieving bonuses that directly involve their own compensation in contrast to hospital administrators. In MIPS, Congress substantially increased potential P4P percentage incremental amounts. However, a series of operational decisions, including exclusion of a substantial number of physician practices from MIPS, have limited bonuses and penalties to far below Congressional intent.

A P4P approach in the United Kingdom, labeled the Quality and Outcomes Framework (QOF), was initiated in 2004 and provided bonuses of more than 25 percent to primary care physicians; yet the approach has elicited mixed reviews about whether the modest quality improvements have been worth the financial investment (Minchin et al. 2018). (We review the QOF in more detail later.)

Despite P4P’s widespread adoption in more than a dozen countries, including the US in Medicare, there is little evidence that P4P for physicians, hospitals, and other providers has improved quality. Further, there has been little assessment of P4P’s effect on non-measured quality, supporting concerns that providers’ attention has been diverted to what is being measured and rewarded, and away from the intrinsic commitment to quality that professionals have.

**Strengths**

- Most payment methods’ primary impact is on volume of services produced. P4P introduces into payment policy emphasis on the quality of care produced, a core element of value that is generally missing in base payment methods.

- P4P permits payers and purchasers to emphasize which aspects of performance deserve priority—for example, an opportune use of P4P to focus on aspects of care that might be compromised under the incentives of particular base payment models (e.g., achievement of preventive service provision in primary care capitation with its incentives for stinting on care).

- P4P is complementary to public reporting of performance; together, public reporting and P4P impart transparency to better hold providers accountable for the large payments they receive, ideally to help consumers and patients make informed choices of provider, and to support quality improvement efforts.
- P4P can complement base payment methods without changing their basic structures. It offers payers the most practical approach to improving value with providers who are unwilling or unable to accept new forms of base payment.

- As an incremental payment method, P4P can be implemented with varying degrees of intensity, consistent with the context of application, the strength of the measures available for the clinical conditions to which it is being applied, or other relevant factors.

- Although there are clear gaps in what is accurately measurable, a commitment to P4P could create momentum to expand measure sets and approaches to achieving greater measurement accuracy.

**Weaknesses**

- P4P introduces significant administrative complexity associated with acquiring data and verifying it for accuracy.

- Behavioral economics suggests that in professions that require high cognitive skill and high intrinsic motivation, associating better performance with financial incentives could be counterproductive because it might compromise the intrinsic, professional commitment to serving their patients and remaining current and skilled in carrying out their professional responsibilities. The result could be that they perform better on measured performance while worse on a much larger amount of unmeasured quality.

- For many health professionals, there are major gaps in what aspects of care are measurable using current data sources. Therefore, erroneous judgments about a clinician's broad-based expertise and performance are likely.

- Most P4P programs tend to concentrate on clinical process measures rather than outcomes, which are the ultimate purpose of health care. Continued efforts to develop a parsimonious set of outcomes-related performance measures have not been successful.

- Apparent improvement in performance may simply reflect more extensive documentation of what was already being done for patients, thus reflecting improvement in reporting rather than actual performance.

- It is administratively easy and practical for payers to base measurement on objective administrative data, usually from claims such as laboratory test results, but that limits the choice of measures. Clinical data from records, self-reporting by providers, or patient-reported
outcomes is more costly to obtain and not necessarily valid and reliable, even (or especially) with wide use of electronic medical records.

- The relatively small incremental reward and penalty payments common in P4P programs may not be sufficient to counter the much stronger incentives in the base payment methods that produce a much larger share of provider payment.

- In the multipayer health care system in the US, different P4P regimes may cause clinician and practice dissonance in responding to different measures, different measurement requirements, and different approaches to rewards and penalties.

- Attainment approaches in public reporting and P4P that compare providers’ performance may not be fair to providers with more challenging patient populations, perhaps because of socioeconomic factors or unmeasured case-mix differences. Yet, improvement approaches compromise the goals of providing consumers with information for choice and for accountability and make it easier for initially poor performers to receive a disproportionate share of rewards for overdue improvement.

- Clinicians might respond to P4P incentives by altering their professional activities so they perform better on the P4P measures, as they sometimes do in public reporting program (e.g., by avoiding high-acuity patients with greater likelihood of experiencing poor outcomes or “unreliable” patients that fail to comply with process measures).

**Shared Savings**

The shared savings method has been introduced partly to give providers an “on-ramp” or “training wheels” for moving away from forms of fee-for-service payment to bear risk. This form of incremental payment, which some consider a form of pay-for-performance focused primarily on spending reductions, is commonly associated with accountable care organizations (ACOs). Yet, the generic approach may be applied to any type of provider organization. In current Medicare and commercial insurance demonstrations, shared savings reward or possibly penalize (shared risk) ACOs for their spending performance in relation to spending targets for a population of patients attributed to the ACO. In the Medicare ACO model, base payments continue using established methods—typically, fee schedules for health professionals and DRGs or per diems for hospitals, while ACO entities can receive additional payment if their spending for beneficiaries is lower than a spending target. When the ACO achieves savings, it can then distribute its share, possibly adjusted by performance on a set of quality
measures, to its constituent members. In this approach the incentives on the individual clinicians, then, is not the same as that on the separate ACO to which the clinician belongs. Put another way, the incentives may not reach down to the level of practicing clinicians.

Shared savings programs can be characterized as “upside-only” (“one-sided”) risk or “upside-and-downside” (“two-sided”) risk, which here we have labeled “shared risk.” Consistent with the training-wheels notion, most ACO arrangements start with upside-only risk and migrate to shared risk over time, once the ACOs have had some experience with shared savings. Some shared savings programs, including Medicare’s, require a “minimum savings threshold,” which separates spending reductions due to successful ACO efforts from spending reductions due to random variations; the more people assigned to the ACO, the lower the threshold for receiving shared savings. Shared risk models generally give ACO providers a larger percentage of savings bonuses in exchange for the financial risk they are willing to assume.

An essential part of the shared savings approach as applied to ACOs is reliance on quality performance measures. An ACO’s performance on quality measures determines how much of a bonus it is eligible to receive for reducing spending against its target. In contrast to many P4P programs that target individual providers, ACO shared savings programs use measures that reflect the quality the ACO provides its patient population.

A fundamental difference between global capitation and shared risk approaches is that the former is a base spending method (not reviewed in this paper) and includes almost all of the revenues a payer provides the recipient organization, whereas the latter provides relatively small, incremental rewards or penalties placed on top of other base payment methods. As a result, the incremental shared savings approach does not offer providers the same opportunity to change their business models and cultures or furnish the same predictable cash flow to support change.

Strengths

- One-sided shared savings does not require provider organizations to take on major financial risk, something many such organizations, especially small organizations lacking a strong financial base, are not in a position to do. It establishes gentler, perhaps more realistic, positive incentives that can provide a reasonable entry for organizations that are new to risk-bearing and lack capital to manage population-based payment methods such as global capitation.
The approach to using historic spending as the base for spending targets may be more practical for many organizations whose spending exceeds the average. Shared savings efforts presumably target “low-hanging fruit”—savings that may be easier to achieve from a high base.

One-sided shared savings can generally be adopted under PPO arrangements legally because of the absence of risk-taking.

Shared risk fundamentally changes volume-based incentives much as capitation does, but may be more practical to implement. Standard cash flow continues under established payment methods, so the ACO need not take on the challenging role of paying its constituent provider members, although it still needs to decide how to share any savings earned with its members.

The central role of population-based, quality metrics provides some assurance that spending reductions will not compromise quality.

Models that attribute individuals to ACOs typically do not lock patients into a particular primary care physician responsible for approving referrals. Most programs either do not limit choice at all (e.g., Medicare) or are placed on a PPO product platform that has gentle benefit design incentives to influence provider choice.

Even short of a fundamental reorientation to providing care, provider organizations under shared savings can adopt relatively straightforward approaches (e.g., improving transitions of care from hospital to community, coordinating care for patients seeing many different providers, adopting evidence-based guidelines).

Weaknesses

The dominant, base payment methods used in shared savings models typically remain activity-based.

Expecting the small incremental incentives placed on a separate or intermediary ACO organization to reduce spending, to counter the volume-inducing incentives of the underlying payment system may be an unrealistic expectation.

Using unadjusted, historic spending to determine spending targets is unfair to organizations that have had above-average performance on spending, as they have less room to achieve additional spending reductions.

Similarly, under shared savings, there is a law of diminishing returns after the "easy savings" have been achieved. The maintenance of volume-based payment models—especially under
one-sided shared savings—could actually make it more difficult to achieve the major changes in providers’ business models and cultures that are the goals of stronger base payment approaches.

- Operationally, determining whether and to what extent savings have actually been attained can be challenging. The Medicare ACO program has been subject to criticism for its retrospective attribution of patients for which the ACO is responsible and for its non-intuitive calculation of shared savings bonuses.

- As with global capitation, ACOs may need to consolidate and integrate to have sufficient size and scale to meet requirements under shared savings and shared risk methods. This may empower organizations to use their newfound organizational clout to negotiate higher prices with commercial insurers for the base payments that determine most of constituent members’ revenue—as well as strengthen their negotiating position with other payers that don’t participate in the shared savings arrangement.

**Lump Sum Payments**

Lump sum payments are payments to providers intended to cover the fixed costs of a practice or to finance a particular capital expenditure, independent of any activity or population served by the clinical or practice. Some of the PCMH demonstrations included lump sum payments as grants to practices to adopt and install medical home capabilities. Actual maintenance of the medical home activities may be paid through ongoing payment for services or the size of the population served. Accordingly, often lump sum payments are one-time, start-up funding for capacity building. However, they can also be recurring payments for special purposes that would not be accomplished by tying the payments to services rendered or the size of the population served, but rather are made to defray specific ongoing costs to the organization.

An ongoing example of major reliance on lump sum, supplemental payments occurs is Medicaid payment to hospitals. For example, Medicaid Disproportionate Share Hospital (DSH) payments are made to hospitals serving high proportions of Medicaid and low-income patients (Cunningham et al. 2016). DSH payments are limited to the cost of inpatient and outpatient services to Medicaid and uninsured patients minus payments received from Medicaid (including supplemental payments) and from uninsured individuals.\(^\text{17}\)
Nationally, all supplemental Medicaid payments combined, some related to services rendered and some not, amount to about half of Medicaid non-managed care payments to hospitals (MACPAC 2017a). Another example is direct graduate medical education payments in Medicare to support the ongoing costs of running hospital-based medical residency programs. Lump sum payments have been less used to pay costs for clinician practices than for hospitals.

Strengths

- The payments are made to accomplish specific objectives to either build infrastructure or to defray ongoing costs that would not be best made in payment for services.
- Lump sum payments are discretionary and, in theory, can be discontinued for cause or when the purpose has been achieved.
- Administratively, a lump sum payment does not require substantial transaction costs, either for billing or for accountability for payments made.

Weaknesses

- It may be politically more difficult to obtain authority to initiate lump sum payments than to provide ongoing payments for patient-related services rendered. To a lesser extent, but as an alternative, capitation payment provides funds prospectively.
- One-time, start-up lump sum payments typically lack accountability for how the funds are used.
- It may be politically difficult to reduce or eliminate ongoing lump sum payments, especially in the absence of well-defined accountability requirements.
- Capacity building tends to be ignored as a priority in producing delivery system change, despite the fact that payment for services tend not to pay these capacity costs.
4. Recent Payment Improvements in the Medicare Physician Fee Schedule to Improve Value

Congress created the Physician Payment Review Commission, a predecessor of the Medicare Payment Review Commission, in 1986 with the express purpose of replacing Medicare’s method of paying physicians based on their historical charges. The Commission determined that a fee schedule built on a cost-based relative value scale, supported by effectiveness and appropriateness research, would produce a service mix of greater value to beneficiaries and the Medicare program, while also redistributing revenues away from procedural specialties toward so-called cognitive specialties, especially primary care physicians, who rely on evaluation and management services for most of their revenues. When Congress in 1989 enacted and moved to the resource-based relative value scale (RBRVS)—adopted and implemented as part of the new Medicare Physician Fee Schedule (MPFS) in 1992—the development was viewed as a likely “game-changer,” comparable to current hopes for value-based payment models.

Earlier, the paper detailed the problems with using fee schedules, including an inherent incentive to increase volume (and costs), non-payment of substantial provision of non-codified activities that primary care clinicians routinely provide, and lack of support for team-based care and collaboration across a range of specialists, facilities, and social service agencies. We will review already adopted improvements to address these and other deficiencies and provide available evidence on the success of the changes.

Addressing Fee Schedule Incentives to Increase Value

Whether fees are based on relative costs or other factors, inevitably, payment rates will be higher than the marginal cost of providing services because of the substantial fixed costs built into each fee, which in turn, provide an incentive to deliver more services. The problem is exacerbated by the multi-payer health care system in the US because Medicare, on its own, seeks to pay for the average costs of services—both fixed and variable costs—to assure that Medicare beneficiaries have good access to primary care and other clinicians. Paying average costs assures that payment is significantly more than the variable costs and so provides a broad, volume-inducing effect.
In moving to a resource-based relative value scale, policymakers thought they were replicating what a hypothetical, well-functioning market with efficient producers would produce (Physician Payment Review Commission 1989). Indeed, paying on the basis of costs incurred for producing thousands of codes, it was thought, would move toward “incentive neutral” fees, reducing the potential conflict between physicians’ financial interests and the patients’ medical needs (Hadley and Berenson 1987; Pauly 1980). Paying based on “costs incurred” was viewed as objective, less easily manipulated, and less subject to distortions than charges determined in markets that do not promote price competition, the situation in most health care provider markets (Berenson and Ginsburg 2019).

With 30 years’ hindsight, the conceptual and operational challenges of estimating resource costs have become apparent. The Medicare Payment Advisory Commission (MedPAC) recently catalogued a number of flaws in how CMS calculates relative values, summarizing a long string of critiques of the overreliance of CMS on the American Medical Association’s (AMA) advisory Relative Value Update committee (RUC), which has been tilted to the views of physicians performing procedures (Medicare Payment Advisory Commission 2019). Of note, however, after years of efforts to increase the relative value of evaluation and management services, including office visits, which are the dominant source of revenues to primary care practices, a survey involving physicians from more than 50 specialties and about 1,500 responses18 within the RUC process will result in a major upward revision in fees for office visits, scheduled to be implemented in the 2021 MPFS.19

CMS has not yet estimated the new fee levels that will be adopted in 2021 because they will need to take into account any additional changes to coding and calculations of practice expenses in the interim. CMS did endorse the RUC recommendations which would produce a 0.28 percent increase in work units for code 99214 and 0.33 increase for 99213, the two most commonly billed office visits codes. Other office visits are also increased substantially but not as much as these two. It should be noted that this survey was fairly unique in the RUC process. Many surveys have few respondents, many of whom are not necessarily objective because they know the results will directly affect the payment rates they will receive (GAO 2015). Furthermore, the RUC has been reluctant to rely on empirical data from other sources about time spent providing services despite the fact that time is the predominant factor that determines the level of work.

Additionally, research produced for CMS has found that a large percentage of visits assumed to be part of a global surgical fee do not actually take place, thereby inflating the relative cost of the large number of major procedures with global periods (Mulcahy, Merrell, and Mehrotra 2020). Even estimates of practice expenses, a major component of relative value costs, has proved difficult because the absence of accurate, granular data from a range of practices. Practice expense unit costs
vary with rates of capacity use, such that average practice expenses, even if accurately computed, may pay too much or too little when converted into standard fees.

In short, current estimates of relative values that form the basis for fees cannot be considered valid and reliable. Other countries do not rely on the faux-objective approach that the US does to set relative value units and fees. Some European countries and Canadian provinces rely on fee schedules negotiated between governments and physician associations, but in contrast to the US, with stronger representation of primary care physicians. Some Canadian provinces have attempted to establish fee levels to achieve target incomes across a range of specialties, paying more or less for the same service to different specialties based on their service mix.

Obtaining More Accurate Estimates of Resource Costs

Within the US construct of relative value-based fees, one can learn from Japan about how to obtain more accurate resource cost information rather than rely on the “bottom-up,” micro-costing approach currently used in Medicare. For example, over a period of about five years, Japan decreased the payment for advanced imaging, including MRIs, by about 50 percent by reducing the fees in annual increments while gauging the volume response; as long as there was no substantial reduction in the volume of imaging being performed, they assumed that the payments continued to exceed marginal costs by enough to assure continued access for patients to the services. The growth rate in scans declined but remained positive (Ikegami 2014).

In the Deficit Reduction Act of 2005 (DRA), Congress, in essence, applied an approach similar to Japan, simply and opportunistically, as a one-year “doc fix” to pay for deferring implementation of the sustainable growth rate formula that would have reduced fee schedule rates for 2006. The law limited payment for advanced imaging services under the MPFS to what Medicare would pay under the outpatient prospective payment system. (For unclear reasons, CMS had actually been paying more in total to independent imaging centers than to hospital outpatient departments, an atypical occurrence.) A follow-up study found substantially reduced payments for these imaging services with continued provision of the imaging tests, but at a modestly decreased rate (GAO 2008).

Reductions in overpriced—in relation to costs—procedures and tests free up relative value units that are redistributed to all other services, with evaluation and management services, which make up 51 percent of total MPFS spending, the major type of service beneficiary (Berenson and Braid-Forbes 2020). Of note, primary care clinicians, including both family physicians and general internists, receive
more than 90 percent of their revenues for evaluation and management services. Using market feedback, as in the approach adopted in Japan and fortuitously in the DRA to adjust downward overpriced services, can supplement efforts by the RUC, as it continues with its flawed approach that depends too much on self-interested surveys by specialists. Although the two examples provided here focus on overpriced imaging services, the same approach—assessing impact on access when fees are altered—could be used to adjust other, likely overvalued, fees to determine a more accurate estimate of resource costs. The approach would not substitute for an improved CMS/RUC process but rather would serve to supplement the micro-costing estimates work and practice expenses that are the core of the RBRVS concept as currently implemented.

Altering Fees to Accomplish Policy Objectives

Another approach that is occasionally adopted to adjust upward underpaid services is to make an affirmative decision to increase payment to achieve policy objectives. For example, the Affordable Care Act included two provisions to adjust upward payment to primary care physicians: a) it provided a 10 percent incentive payment under the Medicare primary care incentive payment (PCIP) for five years and b) it raised the Medicaid primary care payment rates up to at least 100 percent of the Medicare rate for just two years (Mulcahy, Gracner, and Finegold 2018). The purpose of both provisions was to increase willingness of primary care clinicians to provide Medicare and Medicaid services, respectively.

An outcomes analysis on one year of findings immediately after conclusion of the five-year initiative confirmed that the PCIP program had had a modestly positive impact on the availability and use of primary care services in Medicare (Lewin Group 2014). Researchers estimated that the Medicaid bump to Medicaid rates would increase payment rates by nearly 75 percent in aggregate with wide variation across the states (Zuckerman and Goin 2012). Despite this large increase in payment rates, a recent analysis concluded that among more than twenty thousand primary care physicians sampled, the payment increase had no association with primary care physician participation in Medicaid or on Medicaid service volume (Mulcahy, Merrell, and Mehrotra 2020). The authors of both outcome studies point to the fact that these one-time payment supplements were of short duration, with no promise of extension, such that clinicians may have chosen not to change their participation decisions for the short periods.

There have also been recent efforts to provide longer term improvements by making policy judgments to alter fees. Specifically, MedPAC, observing that its fifteen year effort to reform the
CMS/RUC process for correcting distortions in the RBRVS-based MPFS had not succeeded, recommended providing supplemental payments for all outpatient evaluation and management services in order to correct for the “passive devaluation” of these services in the CMS rate-setting approach, resulting largely from the fact that every year new and revised service codes are adopted and paid for, thereby usually automatically devaluing all other services in the fee schedule (Maxwell, Zuckerman, and Berenson 2007; Medicare Payment Advisory Commission 2018). Until recent years, these new and revised codes included very few new evaluation and management codes. MedPAC essentially is proposing a “work around” of the current CMS/RUC process as a permanent, rather than one-time, supplement, to provide needed payment directed mostly to primary care clinicians. This approach resembles the approach that some countries rely on—to focus more on the aggregate impact of the fee schedule on compensation comparisons across specialties, rather than the Medicare approach of trying to micro-manage the intricate details of altering the work and practice expense inputs that comprise RVUs.

Adding Billing Codes Eligible for Payment

Traditionally, CMS had been reluctant to pay for non-visit-based activities that are commonly performed in primary care practices. The consequence was that for decades most codes added for payment every year were for procedures, tests, and imaging – all readily definable services with relatively generous payment levels (Maxwell, Zuckerman, and Berenson 2007). However, starting initially in 2001 with payment for physicians to certify a home health stay, and accelerating as of 2013 with approval of numerous new codes for payment, Medicare has mined the potential for improving the value of the MPFS by covering services that would improve quality and in some cases likely reduce costs. In addition to covering certification of home health and hospice services, added codes have addressed managing care transitions from hospital to the community, managing care for patient with chronic conditions, advance care planning, managing care for patients with cognitive impairment and behavioral health conditions in collaborative arrangements with behavioral health professionals, and providing add-on payments for time spent that last at least 31 minutes, conducted before and/or after direct patient care. A focus in the 2019 and 2020 MPFS updates have been liberalizing payment for a variety of recently developed tele-health codes. We explore tele-health and e-visit codes in depth later.

In the first two years of use beginning in 2016, the advance care planning (ACP) code had slow uptake. In 2016, 1.9 percent of traditional Medicare beneficiaries 65 or older had an ACP visit. That
increased to 2.2 percent for the first three quarters of 2017 and was concentrated among beneficiaries who died within the given year, reaching 3.3 percent in 2016 and 5.8 percent in 2017 (Belanger et al. 2019). The same study also found that despite the fact that many hospice and palliative care specialists could use the codes, only one-third did so. In contrast, most ACP visits were billed by internists (48 percent) and family physicians (27.9 percent) in 2016.

Transitional Care Management and Complex Chronic Care Management Codes Improve Value

To enhance compensation for primary care activities that occur outside of face-to-face visits, Medicare approved paying for transitional care management (TCM) services in 2013 and chronic care management (CCM) services in 2015. TCM is designed to facilitate the transition from hospital to home and involves a dedicated office visit after the hospital discharge as well as additional care coordination. Chronic care management is a comprehensive set of coordination services provided monthly to patients with more than one chronic illness.

Commenting on the TCM in its final rule for the 2020 MPFS, CMS found that “the utilization of TCM is low compared to the number of Medicare discharges and that TCM services reduced readmission rates, lowered mortality, and decreased health care costs. With the added value of this program, the reimbursement rates will be increasing." [emphasis added]. Similarly, clinicians and practice staff perceive that the CCM "encourages adherence to treatment, provides access to care team members, and improves the continuity and coordination of care, as well as satisfaction," that is, supports core attributes of high quality primary care (O’Malley et al. 2017).

Studies looking at outcomes for use of the two sets of codes have shown that both codes have added quality to care for beneficiaries while reducing costs for the program. For the TCM code, for the period July 2016 to March 2018, adjusted total Medicare costs and mortality were higher among those beneficiaries who did not receive TCM services compared to those who did in the 31-60 days following an eligible discharge (Bindman and Cox 2018).

For the CCM code, looking at only one year of results after the code was implemented in 2015, researchers found that per beneficiary per month, expenditures for CCM beneficiaries relative to the comparison group of non-CCM beneficiaries decreased in the 12- and 18-month follow up periods, by $28 in the former period and $74 in the latter (Schurrer et al. 2017). The reduced rate of increase in expenditures was driven by decreases in facility expenditures for inpatient hospital services, skilled
nursing facility services, and outpatient services, with increases for home health and for ambulatory procedures. The results also suggested better management of end-of-life care, with reduced Medicare expenditures among beneficiaries who died in the follow-up period.

There has been concern about low utilization of both codes since their introductions to the MPFS, as a result of documentation burdens and the cost and complexity associated with active billing, in contrast to receiving PMPM payments to support these activities. However, billing for the codes has increased substantially as practices have become accustomed to their availability and the rules governing their use, with some reduction in reporting burden and increases in payment levels. For example, for the TCM use increased from almost 500 thousand services in 2013 to nearly 1.4 million services in 2018. CMS denial of claims dropped from about 37 percent in 2013 to 5 percent in 2018. Of note, this code is predominantly used by primary care clinicians – 82 percent of claims were submitted by primary care practices (Marcotte et al. 2020).

Two studies of the CCM code had divergent findings, yet both showed increased use of the code over its early adoption. One study of the CCM for its two years of initial implementation showed that claims for the CCM doubled from 1.2 percent among beneficiaries to 2.3 percent between 2015 and 2016, with a CCM recipient receiving a median of three months of CCM services in 2016, up from two months in 2015 (Agarwal et al. 2018). The other study, in contrast, found that about 19 percent of beneficiaries had received at least one month of CCM services with the majority of beneficiaries receiving between four and ten months, on average (Schurrer et al. 2017). In this study, primary care physicians billed for 68 percent of CCM claims, with 42 percent of CCM billers being solo clinicians. A more recent study found that for 2016 claims over a fifth of all TCM claims and nearly a quarter of all CCM claims were billed by a practice that was not the beneficiary’s assigned primary care practice (Agarwal et al. 2020).

There has been concern that fee schedule payments for these important care coordination services are relatively low in relation to the costs of providing the services; however, the fees have increased incrementally as reporting requirements have been reduced. Another factor perhaps inhibiting uptake of these codes is that practices have to invest resources, such as by hiring non-clinician staff to support the delivery of these services, meet administrative requirements for billing, and ensure compliance with those requirements (Agarwal et al. 2018). One study modeling CCM estimated that more than 100 patients would need to be consistently rerolled to recoup the salary of a full time registered nurse to provide CCM services (Basu et al. 2015).
Despite these barriers to adoption, uptake is increasing substantially and improving value to beneficiaries and the program while providing enhanced revenue to the practice. Much of the costs of providing these services reflect variable costs for mostly professional activities provided by the primary care team, i.e., they don’t require an infusion of upfront money, perhaps with a lump sum payment for capacity enhancement, as building a full-fledged medical home might. In short, these codes lend themselves well to fee schedule adoption because they can be codified with a reasonably clear-cut definition of the service. The fee for the service now clearly exceeds the costs of billing, while documentation requirements over time have become less burdensome. An added advantage to inclusion in a fee schedule is that clinicians are unlikely to provide these services excessively as with many other fee schedule services. In short, these codes pay for desirable activity but not for excessive volume as other fee schedule services often do.
5. The Unique Challenge of Paying for Telehealth and E-visits

In the decades since the introduction of the MPFS, technology has evolved to facilitate greater interaction between health care clinicians and patients outside of the office setting—beyond telephone calls, physicians can now have virtual face-to-face visits, review patient-submitted images via email or secure online patient portals, remotely monitor patients using connected medical devices, and share health care data in consultation with health care professionals remotely. Adoption of these technologies (broadly referred to as telehealth, though Medicare has historically used a narrower definition) has been slow in health care, in part due to challenges and concerns related to payment. Barriers to use by patients are relatively low compared with office-based visits, so payers are concerned about a potential large increase in spending. The relative value of these services is also harder to define and operationalize in a fee schedule like the MPFS. In contrast to TCM and CCM services, which are carefully defined, low frequency services, for which a substantial payment relative to billing costs is made, many non-visit communications are potentially high frequency activities, lacking standard documentation requirements, with relatively low or even negative payment to cost ratios. In short, as long-term payment policy, they are not readily amenable to payment through a fee schedule.

In the spring of 2020, there was a large spike in the use of telehealth services necessitated by the COVID-19 pandemic. During the pandemic, there has been an unprecedented drop in face-to-face visits for care; a study by researchers at Harvard found nearly a 70 percent drop in in-person visits to ambulatory practices from early March through mid-April 2020 (Mehrotra et al. 2020). Between April and May, in-person ambulatory visits rebounded somewhat but were still 44 percent lower than the mid-February baseline (Mehrotra et al. 2020). In early April, adult primary care practices experienced a 51 percent decline in total visits (including in-person and telemedicine visits) from baseline; in early May, the decline was less—about 25 percent compared with baseline, that nearly 30 percent of all “visits” were provided via telehealth, documenting a dramatic shift to non-face-to-face based communication with patients both as substitutes for face-to-face visits and for frequent communications by phone, email, and virtual visits through audiovisual platforms (Mehrotra et al. 2020).

The relaxing of telehealth rules by payers has facilitated the shift toward non-visit communication. However, many anecdotal reports that standard fee schedule payments, which generally only cover
services that directly replace in-person visits, suggest that the current mode of frequent communication with sick patients cannot be sustained financially in the long-run. In this section, we review current policy to support telehealth and e-visits under Medicare and changes related to the pandemic, challenges that raise similar issues and opportunities that would affect other payers as well.

**Medicare Policy**

The MPFS has historically paid providers for services during face-to-face visits. In recent years, there has been a shift toward paying for care delivery in non-face-to-face encounters. Medicare established the telehealth benefit in the 2001 MPFS as a service covered under Part B, and its primary aim was to expand access to care in rural areas. Spending on telehealth services has increased from about $61,000 in 2001 to nearly $17.6 million in 2015. Medicare structured the benefit around originating health care sites (i.e., not patient homes) where the beneficiary was located and distant sites where consulting providers were located. Both sites had to meet certain criteria for payment, and communication between the sites had to be live and interactive audio and video systems. Transmitting medical information to another physician or clinician who reviews it later—“store and forward” or asynchronous communication of medical information—is only allowed in Alaska or Hawaii federal demonstration programs.

Certain types of health providers were eligible to be originating sites and were also required to be in a county outside of Metropolitan Statistical Area (MSA) or in a rural Health Professional Shortage Area (HPSA) in a rural census tract. Qualifying distant site providers were required to be licensed in the state in which the beneficiary was located. CMS paid originating sites a facility fee (about $26 for HCPCS Q3014 in 2020) and paid distant providers a professional fee for the consultation as if the services were delivered in person in a facility. Medicare beneficiary out of pocket responsibilities are unchanged for telehealth services. A subset of services are eligible for telehealth payments.

Even prior to the COVID-19 Public Health Emergency (PHE), CMS had loosened some restrictions in the telehealth requirements for stroke patients and patients with substance use disorders. The Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act removed the originating site geographic conditions and added an individual’s home as a permissible originating telehealth services site for treatment of a substance use disorder or a co-occurring mental health disorder. Similarly, the Bipartisan Budget Act of 2018 removed the originating site geographic conditions and added eligible originating sites to diagnose,
evaluate, or treat symptoms of an acute stroke. Starting in 2020, Medicare Advantage plans could also offer more generous telehealth benefits as part of their benefit packages.

Restrictions and Challenges

The narrow definition and strict requirements for telehealth services in Medicare has contributed to slow increases in use prior to the pandemic and challenges related to documentation (CMS 2018). An HHS OIG review found that 31 of 100 selected telehealth claims in their analysis did not meet Medicare requirements (Jarmon 2018).

Another challenge for telehealth services is that patients need reliable access to internet service and devices on which to communicate with their health care providers. Many people lack the necessary technology—while almost all American adults have a cell phone, a smaller share (81 percent) own a smartphone and there are wider disparities by race, income, education, and age. Likewise, there are disparities in broadband deployment across the country.

Medicare does not require prior consent to receive telehealth services, but many states may do so within their Medicaid program or as part of regulations for the behavior of health professionals. Different states have different policies, which complicates care delivery that may cross state lines. In addition, written consent may be required. Telehealth poses some additional challenges for consent, such as confidentiality, privacy, and assurance of understanding. As such, the American Medical Association developed a code of ethics for telemedicine.

During the COVID-19 PHE, CMS has used the waiver authority provided in HR 6074, the Coronavirus Preparedness and Response Supplemental Appropriations (CARES) Act to loosen restrictions in telehealth services, announcing an additional round of regulatory changes at the end of April 2020. The site and rural requirements no longer apply, meaning patients in both rural and urban areas in non-health care sites like patient homes can receive eligible telehealth services. CMS is not enforcing the requirement that patients have an existing relationship with the distant site providers during the PHE. Providers may use telephones that have audio and visual capability to communicate with patients, as well as everyday technologies such as FaceTime and Skype.

Before the COVID PHE, CMS did not recognize clinically related phone calls with patients as separately payable despite the presence of CPT codes for them. CMS initially established national payment rates reflecting estimates of work and practice expenses (e.g., about $15 for a 5-to-10-minute phone call (99441) with an independent practice). The April regulatory changes increased
payment for audio-only telephone “consultations” (HCPCS codes 99441-99443) equivalent to payments for office-based visits (99212-99214), retroactive to March 1, and expanded the list of services that can be delivered in an audio-only format. The payment for 99441 increased to about $46 for independent practices. The HHS Office of Inspector General is providing flexibility for providers to reduce or waive cost-sharing requirements during the PHE. The administration also announced a licensure waiver to allow health care clinicians to practice across state lines if they have an equivalent license from another state, but authority and implementation remain unclear. There is an interstate Medical Licensure Compact, of which the majority of states participate.37

Virtual E-check-ins in 2019 PFS

In the 2019 PFS (prior to the pandemic), CMS introduced payment for virtual check-ins. A virtual check-in pays professionals for brief (5-10 min) communications with established patients that mitigate the need for an in-person visit, meaning they do not result from a service within the past 7 days, or result in a service in the next 24 hours, or next available appointment. These visits are distinct from Medicare telehealth visits, which are treated the same as an in person visit and are intended to substitute for it; the distant provider can bill for the telehealth visit using the in-person code for that service.

The Medicare virtual check-in is a specific type of non-office-based visit. Virtual check-ins may be conducted over the telephone, as well as through more advanced technologies such as patient portals and other HIPAA-compliant video communication technologies. While a Medicare telehealth visit is treated and paid as an in-person visit, the virtual check-in is not paid like an in-person E&M visit. The payment rate is about $15 in 2020 (G2012), similar to the initial pandemic rate for 99441. Notably, during the PHE, the payment rate for 99441 is roughly three times higher than for the virtual check-in. CMS also established payment for remote evaluation of recorded video or images submitted by an established patient (G2010) that are not associated with a recent (i.e., within past 7 days) evaluation and management service or upcoming procedure. Outside of Medicare, some insurers may only pay for virtual visits that replace office-based visits. In the MPFS, Medicare pays for remote patient monitoring—in both bundled payments for certain services as well as separate payments.38
Implications for Payment Policy

As noted above, CMS now pays about $15 for a virtual check-in under standard MFPS requirements for determining relative value units (and during the COVID crisis, nearly three times more for basically the same service). It is likely that the billing costs for the practice to seek payment even exceeds the amount received for a virtual check-in, as suggested by a number of studies. For example, a recent study from an academic health center found that estimated processing time and total costs for billing and insurance-related activities were 13 minutes and $20.49 for a primary care visit (Tseng et al. 2018). Because much of the practice costs associated with billing and insurance are fixed, it is unlikely the billing costs for a short telephone call would be much less than this amount. In short, the cost of billing for the virtual check in is more than the $15 the practice receives in the Medicare fee schedule.

As detailed earlier, in the context of the pandemic, Medicare effectively now pays the same rates for telephone-based visits as office-based visits. This emergency measure is an effective tool to quickly push money to financially strapped health care clinicians, but it is not a sustainable long-term approach to payment for telehealth and related non-in-person evaluation and management services that patients have now come to expect and which should be substantially increased as part of comprehensive primary care. The clear implication is that fee schedules are not a good way to support practices actively engaged in communicating with their patients outside of in-person office visits (Berenson and Horvath 2003). When paid in line with underlying resource costs, the transaction costs associated with billing and documentation exceed the return on billing for the service.

However, if fees were substantially higher so that billing for the service would be worthwhile, there would be concern about a proliferation of low cost, routinely overpaid non-visit communications. In contrast to office visits, there would be no time costs to seeking the telephone or email service equivalent to a patient taking substantial time to travel to and from the clinician's office and sitting in the waiting office, a time commitment surely exceeding an hour and often much more. Waiving patient cost-sharing might exacerbate the potential increases in spending, whereas maintaining patient cost-sharing would also generate billing costs far exceeding the trivial cost-sharing amount. In the absence of documentation and oversight, the incentive for practices to bill short, common phone calls as 5-to-10-minute visits is strong. As these calls proliferate, CMS is likely to impose draconian and burdensome documentation requirements.

For these reasons, paying for fee schedule codes for the time spent in much more frequent, but short communications with patients would be flawed payment policy. A policy objective of supporting this type of care, which has become standard with COVID-19, would require a different payment
method that does not pay for actual services rendered to specific patients. Paying for this would likely require a general care management payment akin to capitation, so that the actual services do not have to be allocated to individual patients. Practices would then have control over establishing limits on how available these non-visit communications would be for patients to use. Similarly, an episode-based payment could give practices the flexibility to allocate resources over the course of a particular condition for a specific patient, allowing rapid adaptation to changing care needs.
6. Primary Care in US Public Delivery Systems

Though the majority of the health care system in the United States consists of independent physicians and hospitals who contract with payers like commercial insurance, Medicare, and state Medicaid agencies, there are several areas within our system in which the government plays a much larger role in care delivery. In this section, we briefly describe the financing of public delivery systems, how primary care is delivered, and any recent innovations in primary care delivery.

FQHCs and RHCs

Health centers are community-based and patient-directed organizations that deliver comprehensive, culturally competent, high-quality primary health care services. Health centers also often integrate access to pharmacy, mental health, substance use disorder, and oral health services in areas where economic, geographic, or cultural barriers limit access to affordable health care services. Federally Qualified Health Centers are community-based health care providers that receive funds from the HRSA Health Center Program to provide primary care services in underserved areas. They must meet a stringent set of requirements, including providing care on a sliding fee scale and operating under a governing board that includes patients. Health centers are required by statute to provide health care to all individuals located in the health center’s service area or individuals who are members of the health center’s target population, regardless of their ability to pay. Health centers are also required to be located in geographic areas that have few health care providers or to provide care to populations that are medically underserved (Heisler 2017). HRSA funds nearly 1,400 FQHCs operating approximately 12,000 service delivery sites in every US state, US territory, and the District of Columbia. In 2018, there were more than 236,000 full-time health center providers and staff serving nearly 28.4 million patients. There are also 84 “look alike” health centers that meet the requirements of FQHCs but do not receive HRSA grant funds; these centers served nearly 900,000 patients in 2018.

Health centers receive patient-related revenue such as Medicaid, Medicare, and private insurance payments, as well as federal and non-federal grants (such as Migrant Health Center or Public Housing Primary Care grants from HRSA’s Bureau of Primary Health Care). Medicaid, which represents about 65 percent of patient-related FQHC revenue, for FQHC services, uses either a prospective payment...
system (PPS) or, at state option, an alternative payment methodology to pay FQHCs (20 states do so) (Curt Degenfelder Consulting 2017). The National Association of Community Health Centers encourages additional states to pursue APMs that encompass capitated approaches to FQHC payment in order to better align Medicaid payment with care delivery in health centers, helping ensure that federal grant funds support care for uninsured patients as intended instead of subsidizing low Medicaid payments.

Congress enacted the FQHC PPS to provide payment for the comprehensive services that clinics provide to patients that are not commonly reimbursed in Medicaid FFS, such as case management, translation, transportation, and some mental and behavioral health services. Instead of traditional FFS, the FQHC PPS pays clinics a single bundled rate per patient visit that covers all services and supplies during the visit. The Medicaid FQHC PPS rate was established in 2001 based on historical costs of delivering FQHC services and updated annually (MACPAC 2017b). Starting in 2014, Medicare (which makes up about 12 percent of patient-related revenue) also uses an FQHC PPS. There is a national base payment rate ($173.50 in 2020) that is geographically adjusted and increased for new patient visits, annual wellness visits, and initial preventive physical exams. For comparison, the 2020 Medicare national payment rate for an established outpatient office visit (HCPCS 99214) was $110.43, although in Medicare, practices can bill for services in addition to the office visit.

Rural Health Clinics (RHCs) also provide primary care services to an underserved population, though the eligibility criteria for RHC certification differ from FQHCs and RHCs cannot be FQHCs (HRSA 2006). RHCs do not have minimum service requirements and can focus on one area of care (like pediatrics or OB-GYN). Medicare pays RHCs an all-inclusive rate (AIR) for medically necessary, face-to-face primary health services and qualified preventive health services furnished by an RHC clinician (CMS 2019). RHCs may provide general care management services, and Medicare pays for these using HCPCS codes. Medicare waives the RHC face-to-face services requirement for care management services, and auxiliary personnel may furnish these services under general supervision. RHCs cannot bill for care management services if another clinician or facility already billed them during the same time period. In 2018, there were about 4,500 Medicare-certified RHCs providing care to more than 8 million people in 45 states.

Veterans Health Administration

In the Veterans Health Administration (VHA), primary care is delivered through a patient-centered medical home approach called the Patient Aligned Care Team (PACT). The PACT model promotes
team based, patient-centered care focusing on a personalized, integrated, comprehensive, and coordinated approach to health care. When veterans enroll in the VHA, they are assigned to a PACT that includes a primary care provider (main doctor, nurse practitioner, or physician assistant), a clinical pharmacist, a registered nurse (RN) care manager, and an LPN or medical assistant and clerk (Leung et al. 2019; Nelson et al. 2014; Rosland et al. 2013). “The VA’s funding is appropriated every two years by Congress as a global budget, which is distributed to the VA’s regional networks via a form of capitation that factors in patient demographics, disease severity, and utilization patterns. Network management allocates this funding within the region to pay for staff, facilities, and other resources and is held accountable for performance, as determined by measures that assess quality-of-care, patient satisfaction, and financial efficiency.

The incentives to improve care are both financial and nonfinancial; network leaders and physicians receive performance-based pay and network performance is widely publicized within the system.” Early evaluations of the transition to PACTs in 2010 found modest increases in primary care visits and modest decreases in hospitalizations, with net investments in PACTs exceeding the cost savings of reduced hospitalizations (Hebert et al. 2014; Klein, 2011; Oliver 2007). Staff burnout was reduced and patient satisfaction increased. Another large initiative by the VHA aimed to increase access to mental health services by embedding mental health specialists in primary care clinics (Chang and Simon 2013); research shows this initiative increased access but also increased total spending (Leung et al. 2019).

**Indian Health Service**

The Indian Health Service (IHS) provides comprehensive health care services to roughly 2.6 million American Indians and Alaska Natives (AIANs) from more than 567 federally recognized tribes across the United States. The IHS includes federally run facilities and tribally run facilities, as well as facilities that provide services exclusively to AIAN people in urban settings. There are roughly 335 health centers (both federal and tribal), with a focus on primary care. In 2008, the IHS initiated a patient-centered medical home (PCMH) program called Improving Patient Care (IPC); as part of the IPC, clinics have implemented parts of a PCMH model tailored to fit the needs of the population using IHS services. Clinics received practice transformation support through “change packages” providing a structured approach to implementing changes and technical assistance from improvement support teams. The IHS operates relatively few hospitals, so inpatient care and some specialty care is provided on a contract basis (Timbie et al. 2018).
In addition to clinic-based services, the IHS uses a community health representative (CHR) model to "provide quality health promotion and disease prevention services to American Indian and Alaska Natives through the use of well-trained CHRs, who are integrated into and partner with the patient's health care team. The CHR Program is based on the concept that indigenous community members, trained in the basic skills of health care provision, disease control, and prevention, are best able to gain community acceptance and make use of limited health care resources." CHRs serve as a link between the clinical setting and the community to facilitate access to services and improve the quality and cultural competence of service delivery. They assist by increasing health knowledge of patients and communities through a broad range of activities such as transportation to health visits, outreach, community education, informal counseling, social support, and advocacy. The National IHS CHR Program provides funding, training, and technical assistance to tribal CHR programs to address the health care needs through the provision of community-oriented health services.
7. CMMI and PTAC Primary Care Payment Models

Center for Medicare and Medicaid Innovation (CMMI)

The 2010 Patient Protection and Affordable Care Act (ACA) established the CMMI to test new care delivery models and payment reforms. In the ten years since its formation, CMMI has implemented a number of primary care focused initiatives as well as payment models focused on particular health conditions (e.g. the Oncology Care Model) and comprehensive payment models (e.g. Next Generation ACOs) (Peikes et al. 2020). The primary care-focused initiatives included those focused on particular care settings or populations as well as more comprehensive initiatives. Five CMMI initiatives dealt directly with primary care practices; these initiatives are described below. CMMI also implemented broader initiatives that supported primary care practice transformation: the Health Care Innovation Primary Care Redesign awards (HCIA PCR) that provided grants to test new care delivery models (without payment reform); state innovation models (SIM) supporting statewide healthcare transformation efforts including primary care medical homes, health homes, integrated or accountable care, and payment reform; and the Transforming Clinical Practice Initiative (TCPI), which provided technical assistance to primary care and specialty practices in implementing quality-based approaches without additional payments. CMMI has also announced new primary care-focused payment models slated to begin in January 2021 (more below; Phillips et al. 2019).

The Multipayer Advanced Primary Care Program (MAPCP) was a multi-payer demonstration that began in 2011. MAPCP provided certified medical home practices in eight states with monthly care management fees designed to average $10 or less per patient per month (though this ranged across states), plus additional resources (e.g., quality measure reports, learning collaboratives). The care management fee was intended to cover care coordination, improved access, patient education and other services to support chronically ill patients. Additionally, each participating state had mechanisms to offer APC practices community support and linkages to state health promotion and disease prevention initiatives. By the final year, this demonstration included more than 6,000 providers at more than 800 practices, providing advanced primary care services to more than 3 million individuals, including more than 700,000 Medicare fee-for-service (FFS) beneficiaries. The MAPCP Demonstration infused nearly $125 million in demonstration payments to support the provision of patient-centered
comprehensive, coordinated primary care and enhanced access (RTI International, Urban Institute, and National Academy for State Health Policy 2017).

The experience across the eight states was somewhat mixed, but none were markedly positive. One of the problems observed in some of the states was that commercial insurers dropped out of the coalition, going their own way. Medicare, in this case, actually was a fairly passive participant in most state payer coalitions, with Medicaid and/or commercial plan interests dominating (RTI International, Urban Institute, and National Academy for State Health Policy 2017). Changes in political leadership also contributed to a change in direction of the demonstration in some states, with some loss of interest. The MAPCP final evaluation noted three lessons learned: practice transformation is possible for practices of all sizes, as long as they are provided with sufficient resources, appropriate technical assistance, and aligned incentives and expectations across payers; participation of all payers and alignment of payments are critical; and sufficient time to see results is needed.

The FQHC Advanced Primary Care Practice was a three-year CMMI demonstration project operated in partnership with HRSA that ran from 2011-2014. In this initiative, participating FQHCs were paid a quarterly care management fee of $18 for each eligible Medicare beneficiary receiving primary care to help the FQHC achieve Level 3 PCMH recognition, assist patients with management of chronic conditions, and actively coordinate care. FQHCs were early adopters of the PCMH model (Timbie et al. 2017). The FQHCs also received technical assistance through the program to help support and guide their transformation in all six PCMH standards, as well as to collect data and feedback on their performance. More than 500 FQHCs participated in the demonstration (roughly four percent of the 12,000 FQHC delivery sites). Overall, 70 percent of the 503 participating sites achieved Level 3 status by the end of the demonstration, compared with 11 percent of comparison sites.

The final evaluation report highlights several key findings—care management fees for Medicare beneficiaries, who make up a relatively small share of FQHC patients, were not very strong incentives for transformation; it took practices some time to transition to PCMH Level 3 status; and beneficiaries attributed to NCQA Level 3-recognized FQHCs had significantly better utilization, process, and spending outcomes than did beneficiaries attributed to other FQHCs, with mixed findings on consumer experiences (Kahn et al. 2017).

The Independence at Home (IAH) demonstration project examined whether providing comprehensive primary care services at home for chronically ill patients could improve overall quality of care and quality of life for patients served, while lowering health care costs by forestalling the need
for care in institutional settings. Eligible practices—those with experience providing home-based primary care to patients with multiple chronic conditions and who have at least 200 eligible Medicare beneficiaries—could receive incentive payments if quality metrics and a minimum cost savings threshold was reached. A March 2020 evaluation of the IAH demonstration reports about 8-9,000 beneficiaries received services each year through 14 participating practices; the reduction in average Medicare spending per beneficiary (about $200 PMPM) over the five years was not statistically significant, which may reflect small sample size (Li, Kimmey, and Cheh 2020).

The Comprehensive Primary Care Initiative (CPCi) model was a four-year multi-payer demonstration launched in 2012. CPCi provided population-based care management fees and shared savings opportunities to participating primary care practices to support the provision of a core set of five “comprehensive” primary care functions. CPCi had mixed results on the outcomes assessed in the evaluation and reduced FFS expenditures for attributed beneficiaries, though the decrease was not enough to offset the care management fees (Peikes et al. 2018). CPCi, which is no longer active, was a precursor to the Comprehensive Primary Care Plus (CPC+) program launched in 2017. CPC+ is a national advanced primary care medical home model that aims to strengthen primary care through regionally-based multi-payer payment reform and care delivery transformation (Burton, Berenson, and Zuckerman 2017).

CPC+ includes two primary care practice tracks with incrementally advanced care delivery requirements and payment options. The 3,070 practices participating in CPC+ in 2017 and 2018 were evenly split between the two tracks. There are three major payment elements to CPC+: (1) a risk-adjusted care management fee per beneficiary that is paid quarterly and non-visit-based; (2) performance-based payment incentives that are paid prospectively with retrospective reconciliation, with performance measures including patient experience, clinical quality, and utilization; and (3) payment under the MPFS (Center for Medicare & Medicaid Innovation. 2019). In Track 1, payment continues as usual under the MPFS but in Track 2, fee schedule payments are reduced and shifted into a substantial Comprehensive Primary Care Payment. Practices have options regarding how much of a fee schedule reduction, particularly in the first two years of participation (10/90, 25/75, 40/60, and 65/35 are all options). In years 3-5, which begins with 2020, practices can opt to have either 40 percent or 65 percent of the total base payment, at the choice of the recipient. We discuss design issues in CPC+ Track 2 in more detail on when discussing design issues in hybrid payment models later.

Because CPC+ builds on previous primary care models and was implemented later, only the first annual evaluation report is currently available. The evaluation focused on 2,905 practices (including
13,000 practitioners) that started in CPC+ in 2017. The median care management fees per practitioner in the first year equaled $32,000 in Track 1 and $53,000 in Track 2, with CMS providing the majority of care management fees relative to other participating payers. Other payers also were slower to adopt the reduced fee schedule payments and partially capitated payments in Track 2. Practices reported that a median of 35 percent of active patients were attributed to them through CMS or other payers. In the first year, the majority of Track 2 practices (71 percent) opted for the lowest FFS reduction in exchange for capitated payments, 10 percent; only 16 percent of CPC+ track 2, or roughly 8 percent of all CPC+ participants, opted for a 40 or 65 percent reduction in FFS in exchange for capitated payments. As expected, in its first year, CPC+ had a few very small impacts on quality of care and health care spending, but the care management investments were larger than any cost savings in year 1. Despite the investments in care management, data feedback, and learning support, participating practices also expressed challenges implementing care delivery requirements, though system-owned practices and those with robust health IT infrastructure found it easier to identify the resources for practice transformation and manage reporting requirements.

In addition to initiatives focused on primary care payment models, CMS manages several types of broad Accountable Care Organization (ACO) programs established in the 2010 Patient Protection and Affordable Care Act (ACA) that have implications for primary care payment, some of which are demonstration projects through CMMI. ACOs are groups of doctors, hospitals, and other health care providers, who come together voluntarily to give coordinated high-quality care to their Medicare patients. When an ACO succeeds both in delivering high-quality care and spending health care dollars more wisely, the ACO will share in the savings it achieves for the Medicare program; some tracks include shared risk if costs exceed spending targets. In ACOs, payments to providers within the ACO typically continue on a fee schedule basis, and it is the ACO—a related organization to which providers are affiliated—that assumes the performance risk and associated incentives. Primary care is a central component of these comprehensive ACOs, and organizations differ in the extent to which they emphasize and incorporate primary care.

There has been considerable research on the impact of ACOs, much of which is beyond the scope of this paper. However, related to the role of primary care within ACOs, research has demonstrated that ACOs with a higher share of PCMH practices perform better on cost and quality measures (Jabbarpour et al. 2018). Similarly, physician-led ACOs vs hospitals have greater impact, with greater savings from physician-led ACOs than hospital led ACOs as well as higher levels of measured quality and patient experience (Bleser et al. 2018; McWilliams et al. 2018). The attribution model for Medicare ACOs emphasizes visits with primary care providers, creating an incentive for the ACO to
include primary care physicians and exclude specialists (Lewis et al. 2013). ACOs have different structures for distributing allocating performance-based payments across participants, which can impact compensation of primary care physicians (Siddiqui and Berkowitz 2014).

Lessons Learned from CMMI Experience Thus Far

After 10 years of experience with previous and current CMMI models for all provider types, Peikes and colleagues provide some key lessons for future primary care model development (Peikes et al. 2020). Though practices valued the care delivery innovations, they often struggled to find the time or resources necessary to fully implement desired changes. Busy primary care clinicians need education about what they are required to implement and why, and simplified and harmonized reporting requirements across payers reduce burden on practices. In addition, practices need flexibility to adapt payment models to their circumstances. Those practices with a culture of learning and embracing change seem to have an easier time implementing care delivery changes. Involving providers other than those in primary care can enhance model impact, and the redesign of care can take time to yield impacts.

On a broader level, the review asserted that payment models need to have stronger incentives to counter fee-for-service; multi-payer participation can help achieve this and was generally preferred by practices. Payments need to be clearly defined, relatively simple, and transparent. Data and feedback on performance is necessary, but it must be salient and actionable by practices and training on using data effectively may be necessary for model participants (Peikes et al. 2020).

Recently Proposed CMMI Models

In the summer of 2019, CMMI announced several new models on track to begin in 2021. Primary Care First (PCF) builds on the principles of CPC+, particularly Track 2, by combining capitated and reduced fee schedule payment levels, with greater potential for shared savings and expansion to a larger number of geographic regions. Like CPC+, PCF is a regional multi-payer initiative for advanced primary care practices. The payments in PCF include a professional, risk-adjusted, capitated population-based payment and a flat primary care visit fee, plus performance-based payment incentives that use regional and historical benchmarks for quality and acute hospital utilization. The PMPM payment ranges from $28 to $175, depending on the average Hierarchical Condition Category (HCC) risk score of all attributed patients at the practice. The capitated payment was calibrated to constitute about 60
percent of the total primary care payment amount. The flat primary care visit fee will be about $40 (before copayment and geographic adjustment) and will replace billing for core primary care services; practices may still bill for other services. There are elements of PCF specifically designed for practices caring for patients with complex chronic conditions or advanced illness. Beneficiary attribution will be performed quarterly through a two-step process: voluntary designation supplemented by claims-based alignment. Performance is assessed through comparisons to practice historical performance as well as regional. The evaluation and expansion of PCF will be informative, as the COVID-19 pandemic adds greater urgency to the need for primary care payment reform (Gold, Green, and Westfall 2020).

CMMI also announced Direct Contracting models, which are capitated or partially capitated models that are broad in scope. One of the voluntary risk-based payment arrangements in Direct Contracting—the professional option—includes 50 percent risk for shared savings/losses and a capitated risk-adjusted PMPM payment for enhanced primary care services.

**Physician-Focused Payment Model Technical Advisory Committee (PTAC)**

MACRA was enacted, in part, to improve how the federal Medicare program pays physicians for the care they provide to Medicare beneficiaries. MACRA included a number of provisions focused on expanding value-based payments in Medicare, including the Merit-based Incentive Payment System (MIPS), additional Medicare Alternative Payment Models (APMs), and the creation of the PTAC. MACRA also created incentives for physicians to participate in APMs, and it specifically encouraged the development of certain types of APMs, referred to as physician-focused payment models (PFPMs).

The statutory mission of PTAC is to provide comments and recommendations to the Secretary of HHS on proposals for PFPMs submitted to PTAC by individuals and stakeholder entities. The committee is comprised of 11 individuals with national recognition for their expertise in PFPMs and related delivery of care. The Secretary is required by MACRA to review PTAC’s comments and recommendations on submitted proposals and post a detailed response on the CMS website. PTAC deliberations on proposals occur at public meetings, with the first deliberations occurring in April 2017; as of June 2020, PTAC has deliberated on more than 30 proposed PFPMs that span a number of health conditions and clinical settings, including two focused on general primary care and two on comprehensive care for patients with advanced illness (referred to as AAFP, Dr. Antonucci, CTAC, and PACSSI—see note for full title and submitter information). At this point, none of the proposed
models have been directly implemented by CMMI, but CMMI did state that several models informed the development of new models announced in 2019, including the PCF model set to begin in 2021.57

Broadly, all four of the primary care-focused proposed models submitted to PTAC adopt a partial capitation approach to paying for care-related activities and a pay for performance component to encourage high-quality care.

The PTAC Primary Care–Oriented Proposals

Submitted in April 2017, the AAFP proposal is a partially capitated PMPM payment model with shared risk for cost and quality. AAFP would pay participating APM entities (likely primary care practices) in four ways: a prospective, risk-adjusted primary care PMPM payment for evaluation and management services delivered by the practice; a prospective, risk-adjusted PMPM payment for care management services delivered by the practice; prospectively awarded incentive payments that may have to be repaid based on the practice’s performance; and continued fee schedule billing limited to services not covered in the PMPM payments. In contrast to CPC+, this partial capitation model has a much more restricted role for continued fee schedule payments.

The APM entity could select from two options regarding the PMPM payment for evaluation and management services, one that includes only office-based services and one that includes all services regardless of site of service (e.g., including hospital-based services). The incentive payments would be paid quarterly and reconciled against actual performance annually. The APM entity would select six performance measures, including at least one outcome measure. Failure to meet agreed-upon benchmarks for performance would result in the APM entity having to repay all or part of the incentive payments. The APM entity would also be held accountable for two utilization measures: (1) hospitalization utilization per 1,000 attributed beneficiaries, and (2) emergency department (ED) utilization per 1,000 attributed beneficiaries. Participants would not face full or performance risk for controlling total cost of care. AAFP proposed that the amounts a payer pays for the PMPM and incentive payments should be designed to ensure that total payments to primary care are equal to 12 percent of a payer’s total health care spending on its members. The model would attribute patients to practices primarily through patient choice, with claims-based attribution as an alternative.

The AAFP proposal explicitly references CMMI models in its approach. The proposal abstract frames the proposed model as building on concepts already tested through the CPCi and CPC+. The proposed AAFP model, implemented nationally without a requirement for multi-payer participation,
could enable more primary care practices to participate in an APM. In the proposal, AAFP also discussed the challenge of designed primary care payment models on flawed historical fee schedule relative values for common primary care services. AAFP also based performance-based incentive payments on the CPC+ structure.

Dr. Jean Antonucci, an independent, primary care physician in Maine, submitted “An Innovative Model for Primary Care Office Payment,” to PTAC in March 2018.58 Dr. Antonucci’s proposal is a partially capitated PMPM model with shared risk for cost and quality, though closer to full primary care capitation than the AAFP. Dr. Antonucci’s model would pay physicians risk-stratified, monthly primary care payments per attributed beneficiary in place of payments under the MPFS for evaluation and management services, minor procedures, and office-based tests. The practice would continue to receive fee schedule payments for services where the practice incurs a significant supply cost, such as vaccines and injectable medications over a specified cost threshold. The submitter proposed a $60 PMPM payment for low- and medium-risk patients and a $90 PMPM payment for high-risk patients (based on a self-reported data in the “How’s Your Health” survey).59

Practices would be held accountable for quality based on the results of patient-reported survey data. A portion of the monthly payment (15 percent) would be withheld and paid only if the practice achieved a certain quality benchmark (unspecified in the proposal). The proposed Dr. Antonucci model emphasizes simplicity as a way to encourage provider participation, in contrast to the complexity of existing CMMI models and the AAFP proposal.

In addition to the partial capitation models for general primary care services, PTAC has reviewed two palliative care-focused models that also feature partially capitated PMPM payments with shared risk, the “Advanced Care Model Service Delivery and Advanced Alternative Payment Model,” submitted by the Coalition to Transform Advanced Care (C-TAC), and the “Patient and Caregiver Support for Serious Illness” model submitted by the American Academy of Hospice and Palliative Medicine (AAPHM).

The C-TAC payment model includes wage-adjusted $400 PMPM payments of indefinite duration that would cover palliative care services, care coordination, advance care planning, shared decision making, and 24-7 access to a clinician. Payment would replace the APM entity’s palliative care provider FFS payments for E&M, chronic care management, complex chronic care management, transitional care management, and advance care planning services. APM entities would receive quality bonus payments or shared losses based on the total cost of care for the last 12 months of life.
The AAHPM model proposes two tracks of payment for participating providers. Track 1 includes upside-only payment incentives, while Track 2 includes shared savings and shared risk. Patients would be classified into two tiers based on complexity, moderate and high. In both tracks, practices would receive monthly capitated base payments that would replace payment for E&M services to the patient care team (PCT). In the AAHPM model, Track 1 includes a pay-for-performance model with financial incentives for performance, but it stops short of including insurance risk for total cost of care. In Track 2, however, participating APMs would take on shared risk for total cost of care for a population with advanced illness. PCTs’ eligibility for and extent of savings or risk would be dependent on quality performance.

PTAC provided favorable comments to the Secretary for both proposals, especially because of the need for greater attention to this particular population of very sick patients approaching the end of life. A particular topic that PTAC engaged in, which may help inform payment models for primary care generally, was over the use of total cost of care as the basis for providing shared savings for holding spending below a target amount. In addition to the challenge of accurate risk adjustment for fragile, potentially high cost patients who would benefit from palliative care, members of PTAC raised concerns about the use of total cost of care savings in this patient population, raising concerns about the potential of overt financial conflicts of interest for responsible clinicians who would benefit from withholding costly care, such as for hospitalization, when patients have not indicated they wanted only comfort care. These members of PTAC thought good palliative care would likely reduce care and that direct financial incentives were not appropriate, asserting that utilization measures, such as reduction in categories of avoidable emergency room visits and hospital stays, such as for urinary tract infections, would be more acceptable measures of performance.
8. Selected Primary Care Payment Innovation in Commercial Insurance

In this section, we highlight three of the many commercial insurance-sponsored payment reform models for primary care, the Alternative Quality Contract (AQC) implemented by Blue Cross Blue Shield of Massachusetts (BCBSMA), the Blue Cross Blue Shield of Michigan Commercial Medical Home Insurance model, and the Hawaii Medical Service Association (HMSA) Blue Cross Blue Shield of Hawaii Population-Based Payments for Primary Care. In Appendix B, we included a summary of the U.S. Healthcare primary care capitation model for gatekeeper physicians used in the 1980s and 1990s in HMO products offered by various insurers.

The AQC is one of the largest and longest-running commercial payment innovations, with about 80 percent of providers in the BCBSMA participating in AQC. BCBSMA implemented a global payment approach for providers in 2009 that establishes a global budget for provider organizations to cover all services and costs, including inpatient, outpatient, pharmacy, behavioral health, and other costs and services associated with each of their BCBSMA patients. The initial global budget is based on historical health care cost expenditure levels. It is adjusted each year for inflation and the health status of the provider’s specific BCBSMA patients. The global budget is based on historical levels of health care expenditures and covers all inpatient, outpatient, pharmacy, behavioral health, and other health services BCBSMA patients require. If providers can care for their population for less cost than the budget, they share in the savings they produce; if they go over budget, they share in the excess costs.

This arrangement gives physicians and hospitals the incentive and flexibility to employ novel care delivery models—such as telemedicine, group visits, and follow-up home visits after hospitalizations—that reduce utilization and improve health. Evaluations of the AQC found that compared with similar populations in other states, Massachusetts AQC enrollees had lower spending growth and generally greater quality improvements (Song et al. 2019). Initially, savings did not exceed incentive payments but over time (beginning around year 4), savings accelerated and ultimately exceeded incentive payments. The study found quality improvements in chronic care management, adult preventive care, and pediatric care for AQC participants.

Blue Cross Blue Shield of Michigan (BCBSM) operates one of the largest PCMH programs in the country through its Physician Group Incentive Program. The PCMH program is paired with a PCMH-Neighbor program to enable specialists and sub-specialists, including behavioral health providers, to
collaborate and coordinate with primary care physicians to create highly functioning systems of care. In 2008, BCBSM began its initiative to encourage physician organizations to adopt the capabilities of a PCMH; as of 2018, the program included 1,715 designated PCMH practices, 4,638 designated primary care physicians, and 1.4 million attributed BCBSM members (Blue Cross Blue Shield of Michigan 2019).

Building on the foundation of the PCMH program, BCBSM launched Organized Systems of Care (OSC) to link primary care physicians, specialists, hospitals and other health care partners through the entire health care system. These strategies are integrated into a comprehensive population-based approach to ensure all Blue Cross PPO members receive patient-centered care that provides needed prevention services, chronic care management, and integration of behavioral and medical care. In 2019, BCBSM added a value-based reimbursement component with the aim of holding OSC providers accountable for managing costs. Participants that improve cost performance can earn financial rewards, resulting from the introduction of two-sided risk in subsequent years.

The most recently initiated test of an alternative primary care payment model was undertaken by HSMA in 2016, in collaboration with policy researchers from the University of Pennsylvania. A 2019 peer-reviewed paper described the model and provided first year results of its impact on quality, utilization, and costs (Navathe et al. 2019). This Population-based Payments for Primary Care (3PC) payment model replaced all fee schedule payments with a risk-adjusted payment for attributed members of HSMA insured members, providing 80-90 percent of practice revenue with the rest coming from shared savings and quality metric-derived P4P payments related mostly to performance on quality, rather than cost, metrics. Given the Hawaii physician market, the model had the potential to be effective with large numbers of independent primary care clinicians. Also, it was adopted across all HSMA products, including commercial, Medicare Advantage, and Managed Medicaid. HSMA is the dominant insurer in Hawaii. The stated goal for clinicians to participate was “reducing the pressure for a high number of office visits to generate revenues, to allow greater flexibility for primary care clinicians aimed at population health and quality, not numbers of visits.”

The initial published evaluation relied on a difference-in-difference design comparing the 2012-2015 preintervention period with the 2016 postintervention periods. The intervention group included physicians in the first wave of adoption of the 3PC model, while the comparison group were physicians due to be included in subsequent waves of adoption. The intervention relied on Healthcare Effectiveness Data and Information Set-based metrics that had been in use in the preintervention period, supplemented by eight new measures that had not been included previously. The composite data set are largely screening and prevention measures, such as breast and cervical cancer screening...
and blood pressure control, as well as particular process measures, such as social determinants of health assessment.

The first-year results showed a small improvement in overall quality in year 1, with variable improvement across the measures. Primary care visits declined about 10 percent in both groups, with about a 4 percent greater reduction in the intervention group. There was a modest increase in prescription drug usage. However, there were no statistically significant changes in inpatient and outpatient hospital visits or specialist visits. There was no significant differential change in total cost of care.

It should be emphasized that these were the first-year findings. As with all payment interventions, clinician recipients may need a period of time to adjust to altered incentives to accomplish organizational behavior and cultural change. Another challenge in this study design was that the control group knew they would be included in future waves of adoption and may have anticipated desired practice delivery change. Some findings found desired change in both groups.
9. Primary Care Payment in OECD Countries

Countries use various methods to pay primary care physicians, including fee schedules, capitation, or a mixture of the two, along with other approaches that provide additional payments for performing certain care activities or achieving quality benchmarks. The myriad approaches to pay primary care physicians demonstrate the lack of a one-size-fits-all approach. Table 2 provides an overview, although highly simplified, of the various primary care payment methods used in European countries. Additionally, the table demonstrates that European countries also vary in whether they require primary care physicians to serve as gatekeepers or require patients to register with a primary care physician as a less restrictive form of patient-physician alignment.

In this section, we review a selection of the various international models for the provision and payment of primary care. Specifically, we review OECD countries that are mostly larger in size and where primary payment innovation has been most apparent. These countries have distinctive payment models that draw on the various methods available to pay primary care physicians. Some include variations of those methods that are unique initiatives of the country, such as Denmark's effort to try to appropriately balance the incentives of fee schedules and capitation, while others are some of the largest nation-wide payment reform efforts, such as the United Kingdom's P4P program.

In this international review, we did not include efforts by countries to modify the functioning of their base fee schedule payments. Both Japan and Taiwan are worth briefly mentioning. Japan introduced Continuous Care Fees in its fee schedule in 2014 to encourage physicians to coordinate care for patients with chronic diseases.61 These fees allow physicians to devote time to care coordination, time they otherwise would not have been paid for. Taiwan also implemented a somewhat unique approach in its fee schedule; however, it was targeted toward addressing the underlying incentives to increase volume. Taiwan uses a conversion factor to automatically adjusts fees in the fee schedule down when there are volume increases for those services, an approach similar to Medicare’s troubled history with the sustainable growth rate, but apparently much more successfully.62

The remainder of the section focuses on primary care payment in Australia, Canada, Denmark, France, Germany, The Netherlands, and the United Kingdom, as these OECD countries have experimented with primary care payment beyond fee schedule adjustments.
### TABLE 2
International Comparisons of Primary Care Provision and Payment

<table>
<thead>
<tr>
<th>EU countries</th>
<th>Do PCPs control access to secondary care?</th>
<th>Are patients required/encouraged to register with a PCP or practice?</th>
<th>Primary care payments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>No, no incentive for referral</td>
<td>No, no incentive to register</td>
<td>FFS</td>
</tr>
<tr>
<td>Belgium</td>
<td>No, but incentive for referral</td>
<td>No, but incentive to register</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Yes</td>
<td>No, no incentive to register</td>
<td>Cap/FFS/Other</td>
</tr>
<tr>
<td>Croatia</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Cyprus</td>
<td>No, no incentive for referral</td>
<td>No, no incentive to register</td>
<td>FFS</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>No, no incentive for referral</td>
<td>No, no incentive to register</td>
<td>Cap/FFS/P4P</td>
</tr>
<tr>
<td>Denmark</td>
<td>No, but incentive for referral</td>
<td>No, no incentive to register</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Estonia</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap/FFS/P4P/Other</td>
</tr>
<tr>
<td>Finland</td>
<td>Yes</td>
<td>Yes</td>
<td>Global budget</td>
</tr>
<tr>
<td>France</td>
<td>No, but incentive for referral</td>
<td>No, but incentive to register</td>
<td>FFS/P4P/Other</td>
</tr>
<tr>
<td>Germany</td>
<td>No, no incentive for referral</td>
<td>No, no incentive to register</td>
<td>FFS</td>
</tr>
<tr>
<td>Greece</td>
<td>No, no incentive for referral</td>
<td>No, no incentive to register</td>
<td>Global budget</td>
</tr>
<tr>
<td>Hungary</td>
<td>Yes</td>
<td>No, no incentive to register</td>
<td>Cap/P4P/Global budget</td>
</tr>
<tr>
<td>Ireland</td>
<td>Yes</td>
<td>No, no incentive to register</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Italy</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap</td>
</tr>
<tr>
<td>Latvia</td>
<td>No, but incentive for referral</td>
<td>Yes</td>
<td>FFS/Cap/Fixed pay/P4P</td>
</tr>
<tr>
<td>Lithuania</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap/FFS/P4P/GlobaI budget</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>No, no incentive for referral</td>
<td>No, no incentive to register</td>
<td>FFS/Cap</td>
</tr>
<tr>
<td>Malta</td>
<td>No, but incentive for referral</td>
<td>No, no incentive to register</td>
<td>FFS</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Yes</td>
<td>No, no incentive to register</td>
<td>Cap/FFS/P4P</td>
</tr>
<tr>
<td>Poland</td>
<td>Yes</td>
<td>No, no incentive to register</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Portugal</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap/P4P/Global budget</td>
</tr>
<tr>
<td>Romania</td>
<td>No, but incentive for referral</td>
<td>Yes</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Slovenia</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Sweden</td>
<td>Yes</td>
<td>No, no incentive to register</td>
<td>Cap/FFS</td>
</tr>
<tr>
<td>Spain</td>
<td>Yes</td>
<td>Yes</td>
<td>Cap/P4P/Global budget</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>No, but incentive for referral</td>
<td>Yes</td>
<td>Cap/FFS/Other</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>No, but referrals are most common for accessing secondary care</td>
<td>No, no incentive to register</td>
<td>Cap/FFS/P4P</td>
</tr>
</tbody>
</table>

Australia

Australian primary care physicians are primarily paid based on a fee schedule but receive bonus payments for performing certain activities related to patient care (a form of P4P). Implemented in 1998, the P4P program, called the Practice Incentives Program, provides bonus payments to primary care practices for developing care plans and coordinating care for patients with certain conditions, such as asthma and diabetes, or those needing mental health care (Cashin et al. 2014). These bonus payments account for a little over five percent of federal expenditures on general practitioners. The federal government also funds "Super Clinics" and Primary Health Networks to promote multidisciplinary, team-based care.

Canada

Canada operates a provincial health care model where the provinces establish and negotiate physician fees with the provincial medical association. Primary care physicians are primarily paid based on a fee schedule; however, most provinces provide additional bonus payments or have adopted other payment models for paying primary care physicians. Some fee schedule enhancements include bonuses for complex or chronic disease management or dedicated fees to support the formation of multidisciplinary teams (Joule 2016).

Ontario has been particularly innovative in trying to shift emphasis onto primary care. Ontario officials believed that fee schedules contributed to an excessive focus on acute care and contributed to a shortage of family physicians (Srivastava, Mueller, and Hewlett 2016). Therefore, in the early 2000s, Ontario introduced a new payment effort that replaced fee schedules with a hybrid approach. Physicians could choose among a "menu" of payment combinations, with varying blends of fee schedules, capitation and P4P. The model also encouraged the development of multidisciplinary care teams to better coordinate care (Hutchison and Glazier 2013; Srivastava, Mueller, and Hewlett 2016).

About half of primary care physicians in Ontario are paid through “blended capitation,” where they receive about 70 percent of their income through capitation, 20 percent through fee schedules, and 10 percent more through bonuses. The other half are paid through what they call “enhanced fee schedules,” where they receive about 15 percent of their income through capitation, 80 percent through fee schedules and up to 5 percent through bonuses (Kiran et al. 2015). Accordingly, payments to primary care physicians increased 32 percent between 2006 and 2010 (Hutchison and Glazier 2013).
Denmark

In Denmark, primary care is a focal point of the Danish health care system (Pedersen, Andersen, and Søndergaard 2012). Primary care physicians received higher annual incomes than their specialist counterparts, approximately $146,000 and $133,000 in 2011, respectively. The government intends to attract and retain primary care physicians through higher primary care incomes. About 20 percent of primary care physicians make up the Danish physician workforce (Pedersen, Andersen, and Søndergaard 2012).

Denmark’s payment approach is unique, as policy makers have adopted an explicit approach of balancing incentives to achieve payment “incentive neutrality” with its primary care payment model. The country pays primary care physicians a mixture of fee schedule payments and capitation, representing about 70 percent and 30 percent of their income, respectively. The mix of payments has largely remained unchanged for the past two decades (Forde et al. 2016; Kringos et al. 2015).

The combination attempts to balance the incentive for physicians to under-provide care or refer patients to other providers with the incentive to over-provide care under a fee schedule. Therefore, the fee schedule is used to promote the delivery of certain care activities, such as office or home-based visits, some preventive care, as well as minor surgery, and mitigate the incentive that providers face to refer patients elsewhere (Forde et al. 2016). Capitation, on the other hand, relieves physicians of the incentive to provide unnecessary care for the purpose of additional income. Some question whether the distribution of incentives in Denmark creates the right balance, as fee schedule payments are more dominant and have been for many years (Pedersen, Andersen, and Søndergaard 2012).

France

Prior to 2009, primary care physicians were predominantly paid based on a fee schedule. A shift in focus to emphasize multidisciplinary collaboration and adherence to care protocols to control expenditures led to the development of new payment approaches that provide additional payments to physicians to perform certain care activities or improve care quality.

In 2009, France introduced the “experimentation of new modes of remuneration” (ENMR) to encourage multidisciplinary collaboration. The main goal was to help practices restructure care delivery toward a more integrated approach. Therefore, these additional payments to practices were primarily for care coordination—particularly for patients with chronic conditions—team-based cooperation, and the provision of new services that were not previously paid for (Kringos et al. 2015).
The ENMR began as a pilot program, but the government expanded it across the country in 2011. The ENMR transformed into a country-wide reform in 2015. Currently, more than 300 health groups are participating (Pomey et al. 2019).

France also expanded a P4P program in 2009, called remuneration for public health objectives (ROSP). Direct bonuses are made to individual physicians that emphasize practice reorganization, perform chronic disease management and prevention, and adhere to quality of care guidelines (Cashin et al. 2014). Bonuses are based on the achievement of specific targets, such as physicians’ use of computerized medical charts or electronic claims transmission and the delivery of preventive services or compliance with the established guidelines. In addition to improving quality of care, the ROSP was designed to make primary care more attractive to physicians. By 2011, about a third of primary care physicians had agreed to participate and by 2013, only 5 percent of physicians were not under ROSP contracts (Pomey et al. 2019).

Germany

Most primary care physicians in Germany are self-employed and receive a mixture of an unusual form of capitation and a fee schedule. Both primary care physicians and specialists receive a fee for each patient treated in a given quarter, meaning that they only receive payment for one patient visit per quarter, and only if one visit is made. The payment, thus, is linked to a visit, including telehealth visits, as patients do not have to register with a primary care physician. The quarterly payment amount is adjusted for age and other factors. More time consuming or technical services involve an additional fee. The quarterly payment theoretically creates perverse incentives to withhold subsequent visits in that time period, but also to generate one visit per quarter, although effects on quality, costs, and patient experience with care apparently have not been studied (Kringos et al. 2015).

In 2002, Germany implemented a Disease Management Program (DMP), a pay-for-performance program for chronic illnesses in response to efforts by sickness funds to “cream skim,” or differentially prioritize low-risk patients over the chronically ill (Cashin et al. 2014). Primary care physicians receives annual bonuses for patients that are enrolled in the DMP. Currently, there are DMPs for asthma, breast cancer, chronic heart failure, chronic back pain, COPD, diabetes mellitus Type 1 and Type 2, coronary heart disease, and osteoporosis. There are plans to develop a DMP for rheumatoid arthritis later in 2020. It should be noted that German disease management is based in a general practitioner's practice, in contrast to US models of disease management initiated by private insurers in
which nurses, typically, in call centers unaffiliated with the patient’s primary care clinician engaged patients mostly telephonically under protocols developed by the insurers.

Germany has recently experimented with various integrated care models, including physician-led accountable care organization (ACO), which was piloted in southwestern Germany. Participating providers are still paid based on a fee schedule using the quarterly payment approach but receive additional payments for care coordination and developing care infrastructure, such as health information technology. The ACO has a virtual budget, where they are eligible for shared savings if the total cost of care for the entire insured population in the area is below the budget. Results suggest that the program was successful in keeping costs below the budget (6.6 percent below) and reducing mortality rates for patients participating in the ACO (Srivastava, Mueller, and Hewlett 2016). This approach has not been adopted broadly.

The Netherlands

In the Netherlands, health care is funded by private insurance, but all individuals are required to participate. The federal government allocates a global budget to primary care professionals and sets the general fees and capitation amounts they receive, that can be marginally negotiated with health insurers (Kringos et al. 2015). Primary care payments are a combination of capitation, care coordination fees for physicians and practice nurses, payments for services that either increase practice efficiency or substitute for secondary care, and payments for out-of-hours care.

In 2010, the Netherlands piloted an episode-based payment initiative for diabetes care, care for chronic obstructive pulmonary disease, vascular risk management and Parkinson’s disease. Care groups, consisting of primary care physicians (primarily), nurses, and medical specialists, receive a set payment for delivering all services required to manage a patient’s condition. However, in contrast to episode-based payment being tested in the US, the services covered by the episode payment are routine and pre-determined as part of standard prevention protocols, and do not attempt to include a budgeted amount for all the condition-specific care the patient actually receives or the total cost of care for those patients (Srivastava, Mueller, and Hewlett 2016).

Evaluations of the episode-based payment initiative have had somewhat mixed results. While care groups seem to be conducive to care coordination and adherence to protocols has increased, the reorganization and new payment approaches seem to create additional administrative burdens and contribute to price variation (EXPH 2014).
United Kingdom

In the United Kingdom (UK)’s National Health Service, primary care physicians are predominantly paid via a mix of capitation and fee schedules, although payment approaches often change with change in governments. There are also variations across the constituent entities that comprise the UK. For example, in England, capitation covers essential services and represents the majority of physicians’ incomes, with fee schedule payments made for limited additional services (e.g., vaccinations).72

In 2004, the UK introduced a country-wide P4P program, the Quality and Outcomes Framework (QOF), with the goal to improve investments in primary care practices. The QOF has been the world’s largest P4P program for physicians, applying to GPs, with almost all primary care practices participating and involving evaluation on more than 100 clinically-oriented primary care quality measures. In contrast to most P4P programs, the potential bonuses (there were no penalties) added as much as 25 percent more to of physician practice income (Roland 2004). However, the UK reduced practice bonuses to 15 percent in 2013 because policymakers believed that the higher rate was distorting clinical practice (Roland and Guthrie 2016). The QOF retained in England now awards primary care practices for practice reorganization, patient experience, the provision of additional services, and managing asthma, diabetes and certain other chronic diseases, in addition to a reduced set of prevention-oriented clinical measures (Kringos et al. 2015).

Evaluations of the QOF found some care improvements, but that the program largely fell short in practice redesign and productivity (Roland and Guthrie 2016). Further, some research suggests that primary care practices prioritize QOF-measured activities at the expense of other aspects of care and that the program was associated with physician resentment and burnout (Roland and Campbell 2014). The program has also been criticized for not measuring diagnostic errors and some aspects of appropriateness of care (Cashin et al. 2014). Scotland abolished the QOF in 2016 in favor of other quality improvement models (Roland and Guthrie 2016).

Studies have attempted to estimate the effects of removing the QOF. For example, because the QOF regularly modified the required quality indicators, one study examined the effect of removing 12 quality indicators from the program. The study showed that the removal of these quality incentives led to an immediate decline in quality performance for all 12 in the first year. The changes were small in the years that followed. Specifically, the researchers found that quality reductions were greatest for the indicators related to the provision of “health advice,” but smaller for indicators related to clinical processes or outcomes (Minchin et al. 2018).
10. Review of Core Design Issues in Proposed Fee Schedule/Capitation Hybrid Models

A Historical Perspective on Use of Primary Care Capitation

In recent years, primary care physicians themselves have proposed moving decisively away from fee schedule payments to capitation as the dominant base payment method. The American Academy of Family Practice in 2017 proposed to PTAC a payment model that was essentially primary care capitation, with only a small remnant of fee schedule payments for particular services, with specified approaches to addressing case mix adjustment, performance measurement, attribution of patients to practices, and a series of other design features. A family physician, Jean Antonucci, in 2018 proposed a relatively straightforward form of primary care capitation without fee schedule elements. PTAC found merit in both proposals and recommended to the Secretary of DHHS that their proposals be given serious consideration for development and testing.

CMMI has moved more tentatively toward capitation, initially with a form of per person per month payment labeled a “care management fee” in CPCi while maintaining base payments through the MPFS. However, Track 2 of CPC+, introduced in 2017, moved to near parity between fee schedule and capitation payments. In 2019, CMMI announced a further modification of the hybrid approach in Primary Care First, due to start in 2021.

In short, it seems that capitation, the primary care payment model that was growing and anticipated in the 1990s to become a dominant payment approach, is making a comeback, attaining the policy high ground in many quarters (Zuvekas and Cohen 2010). In considering its adoption now, whether as the clear-cut base payment method or as a prominent component in a fee schedule-capitation hybrid model, it would be useful to briefly review the rise and fall of capitation two decades ago. Earlier we detailed the design, with manifest flaws of primary care capitation models adopted at the height of HMO ascendancy. Here we review retrospective perspectives that discuss primary care capitation within the context of value-based payment reform, more broadly applicable to the range of insurance products, not just HMOs. A review of the capitation history in a Health Affairs article in
2010, as interest in alternative payment models was rising and an essential part of the Affordable Care Act enacted earlier in 2010, observed, “...the local physician and insurance market conditions that created problems for capitation in the past are likely to pose similar challenges to some reforms that are now being proposed...” (Zuvekas and Cohen 2010, 1662).

By 1999, about one-third of physicians had capitation contracts (Dudley and Luft 2001). Even at its peak then, capitation accounted for only 7.4 percent of practice revenues among all US physicians, but higher in primary care specialties, ranging from 12.2 percent of revenues for general internists to 16.4 percent in pediatrics (Zuvekas and Cohen 2010). For those who had any capitation contract, it accounted for 21 percent of revenues at its peak in 1999.

In the 1990s, capitation contracts were more common in commercial insurances products, and somewhat less in Medicaid; in both in the late 1990s, nearly 20 percent of office-based physician visits in commercial insurance and Medicaid. Capitation did not exceed about 10 percent of visits in Medicare, reflecting the fact that the dominant, traditional Medicare only paid physicians by fee schedules. This multi-payer, multi-product distribution and the fact that at its peak, even those with capitation contracts saw only 20 percent of revenues flowing through capitation, represents one of the ongoing challenges with scaling capitation to becoming a dominant payment model for primary care, plausibly too low to generate a fundamental reorientation of practice from volume to value.

Yet, in a multi-payer environment, with traditional Medicare the largest payer, achieving that threshold was difficult even for practices committed to moving off of fee-for-service to capitation. The shift from HMOs to PPOs paying by fee schedules contributed to the entrenchment of fee schedule-based care delivery as usual, even at the peak of capitation in the late 1990s. Professional and global capitation has persisted in California HMOs and in scattered locations elsewhere. Yet, as observed earlier, it was not common for IPAs and medical groups to pass through capitation as the method for paying their constituent clinicians.

When capitation fell, it fell far and fast. By 2010, only 6.6 percent of all physician office visits, as reported in the Medicare Expenditure Panel Survey, were paid by capitation, and by 2013 only 5.3 percent (Zuvekas and Cohen 2010). Because these studies captured the payment approach by which insurers paid practices, and not how the practices paid individual clinicians, the authors of that analysis concluded that in 2013, nearly 95 percent of base payment for office visits was through fee-for-service.

HMOs, but not PPOs and other forms of indemnity insurance, are able to use gatekeeper approaches that requires HMO members for non-emergency care to seek care through their primary
care practice, selecting it from the network of physicians the HMO had contracted with. Receding from memory now, the mid-1900s saw a major consumer backlash to HMOs’ use of prior authorization, gatekeepers, and other utilization management techniques. It is not clear whether capitation per se, outside of the gatekeeper role for primary care practices, contributed to the consumer backlash (Zuvekas and Cohen 2010). (In some HMOs, gatekeeper physicians continued to be paid by fee schedule.) Nevertheless, there was broad concern that pure, primary care capitation presented physicians with a conflict of interest that had to be recognized and somehow addressed, perhaps with robust quality measurement.

The managed care backlash became the dominant theme for a few years in the second half of the 1990s, leading to fierce debate over a Patient’s Bill of Rights. This focused more on managed care restrictions on provider choice and access to services their clinicians were recommending, rather than on how physicians were paid. However, it is important to understand that virtually all the states passed legislation that considered primary care capitation to pose a risk of stinting on care, prohibiting its use outside of closely overseen HMOs (Kongstvedt 2013). Those prohibitions presumably persist.

The decline in use of primary care capitation had causes other than the pervasive and continuing movement of HMOs to PPOs. A fundamental flaw in calculating per member per month rates in primary care payment rates in HMOs was to actuarially convert past fee schedule payments to PMPM rates, without a payment boost that recognized the significantly greater responsibilities and expectations that capitated physicians had as gatekeepers and case managers (Goroll et al. 2007). Many physician practices grew disenchanted with the payment levels they were offered, even if they liked the autonomy and payment predictability that per capita payment allowed them. Some practices also rejected capitation contracts because of the general absence of case-mix adjustment of capitation rates.

By the late 1990s, HMOs had determined that primary care capitation was operationally challenging to administer, especially to improve the problems that physician practices complained about. There was also a growing perception that HMOs were not reducing spending using this payment method; many HMOs abandoned the approach, by the 2000s reverting to the same fee schedule method they were still using to pay specialists. The lack of success again plausibly relates to the multi-payer, multi-product nature of US health insurance. In most markets, only a relatively small percentage of physician revenue came from capitation, and thus the aggregate all-payer payment incentives were too diffuse to induce changes in practice style which the HMOs had hoped capitation would produce (Zuvekas and Cohen 2010). In addition, when many HMOs stopped using primary care clinicians as gatekeepers in response to the managed care backlash, making clinicians accountable for
downstream insurance risk which most had adopted using risk pools, rather than just performance risk of their own services, the process became untenable (Zuvekas and Cohen 2010). Finally, technical issues that affected the feasibility of continuing with primary care capitation would still be applicable today. We turn to design issues now.

Design Issues

Nevertheless, interest in primary care capitation as a primary care payment model has developed in recent years. The payment model that has achieved particular attention in CMMI and among other proponents is a hybrid model that provides payments both through reduced fee schedule payments and per person per month payments, whether the latter are labeled care management or capitation. This is the model that Denmark has successfully used for a few decades and is now the basic payment model for GPs in the Netherlands. The Milbank Fund has called for prompt and decisive adoption of Primary Care First as the most promising value-based model for primary care to be tested.

There are numerous issues in designing a successful hybrid payment model that require full consideration. A full exploration is well beyond the scope of this paper; CMMI’s 2019 report “CPC+ Payment and Attribution Methodologies for Program Year 2020” is 108 pages long. Nevertheless, here we provide some general considerations in thinking about the major design issues and point to some experience and references that might be useful in pursuing these design challenges.

Generosity of Payment

There are many grounds for concluding that payment for primary care physicians, and possibly nurse practitioners and physician assistants, is inappropriately low in comparison to the underlying resources needed to meet the increasing expectations of enhanced primary care practices. Another purpose of payment is to provide appropriate financial incentives for the purpose of addressing the apparent shortage of primary care by making primary care careers more attractive for graduates of health professions schools. Further, there is substantial documentation that the Medicare Physician Fee Schedule, which not only is the fee schedule for Medicare, the largest payer, but also the model that other public and private payers adopt, contains distorted rates in relation to underlying costs. These distortions relatively underpay specialties that mostly provide evaluation and management services, especially primary care physicians, who receive more than 90 percent of their revenues from such services (Berenson and Goodson 2016; Medicare Payment Advisory Commission 2018). Research has
shown that these distortions produce take-home pay that provides a range of procedure- and imaging-oriented specialties more than two times more per hour worked than that received by primary care and other "cognitive" specialties (Berenson et al. 2010).

Further, as documented in Appendix B, a number of studies demonstrate that 25 percent or more of time primary care physicians spend is not paid at all because the activities are not codified and/or approved for payment. Similar data from specialty practices is not available. Finally, results of efforts to increase the percentage of the health dollar directed to primary care, such as in Rhode Island, have demonstrated that important outcomes improve with greater investment in primary care (Baum et al. 2019).

Nevertheless, as we demonstrated earlier, it must be remembered that US physicians are the highest compensated physicians in the world, as measured by their payment compared to the average payment in their country. The implication is that increases in payment generosity to support high quality primary care should include redistribution, not just from particularly highly paid specialties, but also from other sectors of the health care system, especially hospitals. Hospital consolidation over the past two decades has produced health care systems with substantial market power to raise commercial prices far faster than underlying inflation in hospital costs and has resulted in non-profit hospitals accumulating substantial levels of cash and investments, with levels that far exceed the level of community benefits that most non-profit health care systems provide.

The “Right” Mix of Fee Schedule and Capitation Payments

Whereas some policy advocates, including those submitting the AAFP proposal, advocate a wholesale move toward PPPM payments, with only a small level retained for particular fee schedule payments, not including office visits, when given the choice in CPC+, most practices themselves have opted to maintain full fee schedule payments or only a 10 percent reduction in fee schedule levels, as documented in the earlier CPC+ summary, perhaps partly because of reluctance to change what they know for the unknown, even if theoretically desirable.

There are a few ways to think about the proper mix. Denmark explicitly attempts to arrive at a mix that makes the clinician incentive neutral at the margin so that they can act in the best interest of the patient in deciding whether to provide more services. Technically, that would produce a mix in which the already busy clinician would be financially indifferent in deciding whether to recommend that a patient come in for an office visit (or in a new world have a fee schedule-paid phone call or tele-health...
Denmark pays 70 percent base payment through fee schedule and 30 percent through capitation, with some additional P4P bonuses possible.

A second way to think about the best mix would be the percentage that allows the practice to move its business model and office culture from volume to value. This is a difficult calculus because that “tipping point” may vary by clinician and practice. We have not found an empirical basis for determining the tipping point. One might ask clinicians where the behavior change tipping point is with the view that the reluctance of physicians to sign up for much capitation when given the opportunity in CPC+ may mostly reflect caution about operational uncertainty about how CMMI would implement the new approach.

**A Reformed Fee Schedule in the Blend?**

Blended models for Medicare as a possible core of a multi-payer approach typically assume the fee schedule component of the blend would be the standard MPFS. Yet, as demonstrated earlier, the MPFS can be improved to improve value. Arguably, additional changes and enhancements would be helpful in a blend with capitation specifically designed to replace reliance on a pure fee schedule payment method. Codes could be added to promote more collaboration and consultation with specialists in co-managing patients with chronic conditions, to resolve challenges in determining accurate patient diagnoses, and managing polypharmacy, as examples. Finally, fees could be altered to accomplish policy objectives, rather than only relying on determination of relative costs in establishing relative values and fees.

In particular, in the context of alternative payment models, ongoing efforts, heightened by current COVID-19-inspired movement to “virtual” visits of various kinds, to provide more interactions between clinicians and patients outside of in-person visits, generally have not considered fee schedule initiatives to improve generalist to specialist clinician communications. Coding and fee level changes could be accomplished, budget neutral, across specialty care by reducing empirically-determined overpayments for many procedures and test interpretations to permit increases in such communication and collaboration activities. The main point is that fee schedule payment will remain a dominant component in many blend models and could be improved to better promote the value that the blended model is attempting to achieve.
The Approach to Risk Adjustment

Some hybrid payment models, including CPC+, adopt the standard HCC model of risk adjustment that was first adopted for full risk Medicare Advantage plans. However, HCC was adopted to adjust payment for full insurance risk with a dominant factor of predicting the costs related to hospitalizations. This risk adjustment tool makes sense if the primary care practice were held accountable for total cost of care, as in the early HMO capitation models and in recently articulated models, such as that tested by Hawaii Blue Cross Blue Shield.

However, some models do not seek to hold primary care clinicians, who currently generate less than 8 percent of overall health care spending and 2-3 percent of Medicare spending, responsible for addressing total cost of care, whether through actual global capitation or through shared risk based on targets of total costs. An HCC-based approach that predicts total costs of care, emphasizing costs of hospitalization, may not be the best measure of the relative effort and practice-based resources expenditures that a primary care clinician would expend managing patients with different complexities. An alternative view is that primary care capitation should seek to adjust for “primary care activity level,” that is, the care that primary care clinicians should provide, accounting for the great variation between healthy and highly complex patients (Ash and Ellis 2012).

Another set of authors asserted that current methods of risk adjustment relying in diagnosis- and utilization-based algorithms to predict future utilization and costs miss clinical characteristics not present in billing data and may not capture non-clinical contributors to patient complexity (Hong et al. 2015), which could now include social determinants of health. They emphasized that patients identified as complex by their method had only modest overlap with patients identified using claims data, because clinicians consider medical, behavioral, and socioeconomic complexity domains when identifying complex patients. In preliminary research, researchers found that systematic ratings of complexity were a stronger predictor of future emergency department use compared to established claims-based methods, but were worse predictors of future admissions. In this initial study, this approach of clinician estimates of complexity not only predicted poor clinical outcomes and some utilization metrics, but also was a good predictor of the “burden” on primary care teams, perhaps the most important need when using capitation primarily in a payment model.

In short, as a payment method, risk adjustment might address performance risk—the time and effort clinicians themselves provide for patients of different complexity. Full insurance risk would seem appropriate for other forms of capitation for large organizations, including ACOs expected to reduce total costs of care. Other clinically-related risk adjustment approaches have been proposed in...
the context of adjusting for performance risk when capitation is used as a payment method, including ones that rely mostly on patient surveys assessing their own health status, as proposed in the straightforward Dr. Antonucci proposal’s risk adjustment method.

Perhaps even more importantly, risk adjustment for a primary care capitation as a full replacement for fee schedule payment for office visits needs to be a much more accurate predictor of performance risk than in payment models that continue to include a substantial percentage of fee schedule payments; fee schedules serve as a reasonably effective form of risk adjustment for most patients. More complex patients typically generate many visits, whereas healthy patients do not. In short, a near 50:50 fee schedule: capitation blend permits the risk adjuster to be more clinically oriented, rather than solely claims based, and less complex.

**Attribution/Alignment of Patients to Practices**

It would be reasonably straight-forward to use volume-based payment principles to make lump sum payments to practices to cover the costs of a range of activities that are amenable to fee schedule payment, such as the range of short phone calls and emails that clinicians and their staff make or should make to patients and patient families, specialists, pharmacies, and social service agencies, and to support medical home activities, such as team meetings. Such an approach could, for example, provide a monthly payment as a percentage add-on to practice billings to the paper. Unfortunately, such an approach would reward volume of fee schedule services and likely freeze innovation.

For this reason, it is desirable to associate or align patients to their primary care practice to base the financial support for these activities on the population served rather than the volume of services rendered. Making sure the prospective payment for each month’s capitation goes to the correct practice when a patient retains full freedom of choice of provider, as in traditional Medicare, Medicare Advantage and commercial PPOs, and in many Medicaid programs, is necessary—and challenging. CMMI has had to address this issue for ACOs as well as primary care models relying on monthly care management fees or more robust capitation payments. Since attribution was first adopted for ACOs, CMMI and policy opinion, such as MedPAC’s, have encouraged movement from passive attribution to assigning individuals directly to practices based on retrospective claims review, to an approach in which patients designate their practice “home,” which for many reasons may frequently change during the year. Voluntary alignment to determine who gets the capitation payment seems more compatible with the goals of high quality primary care than passive assignment, retrospectively based on claims analysis.
Currently, CPC+ encourages beneficiaries to attest to the health care practitioner and practice that they consider to be responsible for providing and coordinating their health care (Center for Medicare & Medicaid Innovation 2019). In this way, practices that qualify as eligible practitioners under CPC+ rules can be encouraged to engage beneficiaries, attempting to explain the value to seeing the practice as their medical home, available to help them navigate through the complex health care system and as the core coordinator of their care. CPC+ has a process in place for beneficiaries not attributed through voluntary alignment, to align them for payment purposes based on a review of 24 months of claims in which CMS attempts to ascertain where these beneficiaries received one or more primary care visits, with an initial priority for Annual Wellness visits and Welcome to Medicare visits and then, in the plurality of eligible visits. We are not aware of data demonstrating how many choose the preferred approach of voluntary practice designation.

In designing attribution for payment of the capitation payment, it would be important to explore how CMS and other payers have attempted to assure meaningful voluntary patient decisions on identifying their medical home, when it is not an obligatory requirement for an HMO product. Again, this design issue is one where having a blend makes the design feature easier. If payments were all by capitation, except perhaps for limited services such as immunizations, as some have proposed, there would not be claims data on which to make retrospective claims-based alignment. Under pure primary care capitation, payers may require practices to produce encounter data for various purposes, including risk adjustment or performance assessment, as well as for alignment that determines what entity receives the monthly capitation payment, thus undermining the argument that pure capitation produces much lower administrative burden on practices. Yet, experience has demonstrated that encounter data that is not directly connected to payment lacks the reliability that claims for payment provide.

**Accountability and Performance Measurement**

CPC+ specifically rejected restraining total cost of care as an expectation of participating practices with the explanation that primary care practices lack the controls over how and where patients receive care to actually be held accountable for spending; rather, CPC+ holds practices accountable for quality and service use. The program measures quality by Patient Experience of Care surveys and electronic Clinical Quality Measures, and utilization performance through measures of hospital and emergency department utilization. A central design decision around accountability relates to the issue of measuring total cost of care. As noted above, the decision on whether to hold practices accountable for total spending helps determine the kind of risk adjustment that would be most appropriate.
Performance measurement has been subject to extensive discussion for years, with growing frustration that the desirable aspiration of measuring performance for accountability and for consumer choice has gotten bogged down in substantial administrative burden, with attendant high costs, while distracting clinicians from their core, patient care responsibilities. The search continues for the holy grail of core measures to be used for various purposes. This paper will not engage directly in that discussion.

Nevertheless, there may be particular aspects of primary care and a hybrid fee schedule-capitation payment model that suggest need for particular approaches to performance measurement, whether for public reporting or as part of pay-for-performance regimes. The Hawaii demonstration used a set of mostly primary and secondary prevention measures, among the easiest and most accurate ones to measure because most can be obtained from claims; these prevention-oriented measures are core responsibilities of primary care. Those measures do partly address concerns that capitation can lead to stinting on services. Indeed, even in a hybrid model that attempts to balance fee schedule and capitation payments, practices can still accept the monthly payment and not provide the commensurate level services. Measuring prevention services helps, but does not assure, that other forms of stinting are not occurring, forms that are not easy to measure. Patient experience with care through surveys provide some level of ability to detect stinting and could be a basic component of measurement attempting to uncover stinting.

Phillips, Bazemore, and colleagues recently have worked to develop measures of the core attributes of primary care—the 4Cs of Starfield, namely first Contact, Continuity, Comprehensiveness, and Coordination (Bazemore et al. 2018; Bazemore et al. 2015; Etz et al. 2019; Starfield, Shi, and Macinko 2005). Early research has shown that measures of comprehensiveness and continuity, as they define, are associated with lower costs and hospitalization. Additional validation of these measures would provide quality measurement more specific to the vision of high quality primary care than a generic measurement package and could be combined with the kind of utilization measures used in CPC+, along with patient experience measures, to provide a composite measure set for the hybrid approach. Arguably, doing well on this measure set would address total cost of care, without specifically providing financial incentives for primary care to be accountable costs. In addition to measuring utilization of avoidable hospitalizations and emergency department visits, the package could also measure referral rates to specialists, which have been burgeoning over the past two decades (Barnett, Song, and Landon 2012). Referral rate levels would be relevant to measuring practices for continuity and comprehensiveness and important to mitigating the incentive in capitation to refer liberally (Song, Sequist, and Barnett 2014).
11. Conclusion

The COVID-19 crisis has created urgency for moving more decisively away from continued reliance on fee schedule payments as the base payment method for primary care, given the challenges of paying for a myriad number of virtual visits, phone calls, and emails through a fee schedule mechanism. The widespread public and professional acceptance of communications that do not require in-person office visits creates a demand for a payment model that can better accommodate this change in practice without producing unsustainable fee-for-service volume. The need to support non-visit-based communication raises the profile of other objectives for redesigned primary care practice that also are best supported through per capita payments not tied to specific services provided.

Some propose moving almost totally away from fee schedule payments to pure primary care capitation. In theory, a strong case can be made for payment that allows clinicians maximum flexibility to care for their patients without the constraints of fee schedule payments that do not pay for many activities that primary care clinicians already carry out for their patients. Nevertheless, while perhaps aspirationally desirable, pure primary care capitation raises formidable operational challenges, especially when used in the predominant insurance products that permit patients freedom of choice of provider.

A further challenge is that so far, in CMMI demonstrations, clinicians with the choice have largely been reluctant to select more than nominal per capita, monthly payments, perhaps as much out of concern about the unknown as for affirmatively preferring fee schedule payment. And these were clinicians willing to participate in a demonstration that would change off of standard fee schedule payments; presumably others were even less interested in moving away from what they know: fee schedule payment. Current cash flow problems during the COVID-19 crisis may move clinicians to a more positive appreciation of the benefits of reliable, monthly payments independent of services provided.

Design challenges of pure capitation are less, albeit still, present with hybrid payment models that provide a roughly equal balance between fee schedule and per capita payments. For that reason, we have reviewed a number of the core design issues that a blend model would have to grapple with, many of which that CMMI have already engaged in CPC+ and in 2021 in PCF. A successful model, with lessons learned about design, might become the “right” model for primary care, or could serve as an interim approach for moving to a purer primary care capitation model, perhaps for those practices that have earned the trust of providing high quality and cost-effective care through this payment
model that grants clinicians the greatest autonomy and flexibility in deciding how best to care for their aligned patients.

Finally, although primary care capitation to an individual clinician in a solo or small practice or in a multi-specialty group practice offers the possibly of constituting part of a value-based payment model, it should not be confused with professional or global capitation to a large organization that bears substantial financial risk and is held accountable for the quality provided by the organization as a whole. One of the core differences is that it is reasonable to expect the primary care practice to bear performance risk for their own time and efforts but not for insurance risk related to total cost of care. If successful, as proponents hope, a blended fee schedule-capitation model might reduce the trajectory of health care spending without giving practices strong financial incentives tied to their performance on reducing total costs.

In short, the main focus of primary care payment reform should be centered on creating a sustainable payment model to encourage more physicians to become practicing primary care physicians, at the same time providing expanded access to growing patient populations needing high-quality primary care, and not to heroic, but unrealistic, expectations that the primary care workforce should have strong incentives to decisively bend the curve of total health care spending.
Appendix A. Issues in Creating a Logical Typology of Primary Care Payment Models

There have been a number of provider payment typologies developed, with a few focused on physician payment in general and primary care clinicians in particular. The various typologies classify payment methods focusing on different dimensions of payment, such as the type of provider (hospital, physician, etc.), risk-based or not, individual clinicians or institution, unit of payment (per service, per episode, etc.), or based on a continuum of “value” (Berenson et al. 2016). We propose to use and briefly discuss the following dimensions that we are using to classify primary care clinician payment methods to explain the basis for which payment methods we are considering:

- Base versus incremental payments
- The unit of payment
- The recipient of the payment
- Fixed total versus activity-based payment

Base versus Incremental Payments

An important distinction in classifying payment methods is whether the payment method represents a base approach in which the majority of revenues derive from payment or an incremental approach in which an incremental payment of performance bonuses or penalties, or additional payments for specific purposes is provided on top of the base payment. Until recent years, payment tended to simply accept the base payment methods and then focused on additional methods to influence behavior at the margin, through P4P and shared savings approaches. Yet, in many cases, the marginal incentive with incremental payment pales next to the basic payment incentive, often, to generate more volume.

A prime example is the limited success of the hospital readmission penalties to reduce readmissions—a penalty that in dollar terms would be less than the loss of DRG-based revenue hospitals would suffer if they successfully reduced readmissions and, likely, subsequent admissions
through effective programs educating patients about when and how to seek hospital care (Coleman et al. 2006). Similarly, it is unlikely that a small, few percentage points, incentive through P4P to reduce unnecessary services would have much impact on behavior if layered on a base fee schedule payment method that quite generously rewards those services.

The difference between the base payment method in use and the incremental ones is usually easy to distinguish, e.g., Merit Incentive Payment System (MIPS) payments layered on top of standard fee schedule payments. Or a case management payment on top of fee schedule payment as was common in PCMH pilot programs in the early part of the last decade (Bitton, Martin, and Landon 2010).

However, the evolution of hybrid or blended base payment models will reduce the salience of distinguishing between base and incremental payment methods. The paper explores in detail different approaches for determining the relative weights of different payment methods in hybrid payment models.

The Unit of Payment

There are three fundamentally different units of payment that can be used to classify payment methods for clinician patient care activities: time commitment-based, service-based, and population-based. Salary is the prototype of a time commitment-based payment unit—the health professional is paid for her time commitment, which can be as short as a session or as long as an annual salary, classically independent of the activities the professional actually performs or the number of patients under her care (Berenson and Lazaroff 2019).

Service-based remuneration depends on activity—the specific services provided and recognized for payment. Policy shorthand often asserts that under fee-for-service providers get paid for each service they provide, needed or not. This is not accurate and is the primary reason we do not use the term fee-for-service in describing a base payment method. Fee schedules only pay for services that are codified and that the payer determines are actually covered for payment. Research has shown that at least twenty-five percent of the time primary care physicians spend providing patient-related activities is not paid under standard fee schedules (Baron 2010; Chen et al. 2011; Farber, Siu, and Bloom 2007).

Population-based payment varies as a function of the size of the population the provider serves, regardless of the nature and level of activity by the health professional or practice receiving the payment. Capitation—payment per capita—is the classic form of population-based payment. Note that
under salary and population-based payment, payment encompasses all activities the clinician carries out as part of patient care, whereas service-based payment only considers those activities deemed eligible for payment. Payment levels for the defined services in fee schedules can attempt to add in payment for activities not codified for payment but are associated with paid services, for example, in the MPFS, pre- and post-visit time attendant to an office visit.

The CMS-sponsored Learning Action Network classification system considers condition-based episode payment to be population-based (Health Care Payment Learning and Action Network 2017). However, in contrast to capitation, condition-based episodes are made for particular patients, whereas true capitation provides aggregate payments and allows the payment recipient to allocate costs as it determines no matter that the aggregate payment was based on average or expected costs for a population of individuals with specific characteristics.

It gets more complicated. Even prototypical pure payment methods like fee schedules can contain elements of payment based on other units of payment. Payment for office visit codes in a fee schedule can be based on time spent during the visit. In the MPFS, renal physicians are paid a monthly amount for caring for patients with end-stage renal disease, a classic form of capitation, now modified to have different monthly fees for monthly care for renal disease that are modified based on the number of visits in the month.78 Similarly, all routine visits provided by the operating surgeon occurring within 90 days after a major surgery are not separately paid but are rather included in a global payment, a form of episode-based payment that lies somewhere between service-based and population-based payment.

The Provider Recipient of Payment

Payment typologies often sort payment methods by the type of provider typically receiving the payment. This serves two basic purposes: It provides a practical source of differentiation, in that one can readily look for all potential payments to hospitals or to an intermediary organization, such as a medical group which then is responsible for distributing the funds to their constituent members. The clearest example of a payment method available to intermediary organizations but not to payers is health professionals’ salaries. Currently, more than half of practicing US physicians are compensated primarily by salary, but the source of payment is the entity that employs them, not the third-party payer (Boukus, Cassil, and O'Malley 2009). Indeed, in the absence of a single payer, such as the Veteran’s Administration health system, it is virtually impossible for different payers in a multiple-payer health care system to use salary as a primary method to directly pay for physicians’ services.
The distinction between the payer, the intermediary organization, and the provider of services is important and particularly relevant with some newer payment models, such as shared savings. Such models directly maintain payer-generated payment flows to providers and also might maintain separate payment flows to the intermediary or separate ACO. Many health care systems allocate resources for a population using one payment mechanism (macro-level), while compensation for care within the organization employs a different mechanism (micro-level). Further complicating matters, a constituent member can be a medical practice, not the individual physicians within the practice, creating yet another tier of payer and payment recipient. In sum, there may be little relation between the incentives embedded in a payer’s payment method and the incentives the service provider actually receives after the payment is dispersed to members of the organization.

One final classification issue is raised by the number of tiers of payment recipients. For example, in many European countries and increasingly in the United States, physicians are hospital employees. It is common in Europe and some other OECD countries to “bundle” facility and professional services into a single payment for all services provided. But some payment models, as is typical in the US, continue to make separate payments for professional and facility services, even when health professionals are hospital employees.

Whether the fees are actually combined with the hospital payment or not, this approach might properly be labeled a bundled payment. Using provider type as a sorting parameter has appeal because it is relevant to how the health care system is organized. Yet, the approach tends to freeze in place organizational distinctions, which is inconsistent with fostering evolution of payment and delivery models. Increasingly, new payment methods are trying to break down organizational silos and replacing them with various forms of integrated care. Much of the activity in bundling services across provider types to promote integration cannot be well captured in a typology organized through classic differentiation of provider types—health professionals, hospitals, ambulatory facilities, and so on. Alternative classification approaches might focus more on incentives in the payment methods than on the provider recipient.

Fixed-Total versus Activity-Based Payment

Another frame for describing payment models is based on whether providers receive additional revenues when they provide additional services. The practice of health care, like other industries, has fixed and variable costs associated with the delivery of services. In purely activity-based payment approaches, payments should be sufficient to cover both fixed and variable costs. Further, as long as
the payment exceeds the variable cost of production, providers have incentive to produce additional services, also creating the risk of overprovision. To counteract this incentive, activity-based payment approaches are sometimes supplemented by lump-sum payments to cover fixed costs, so that activity-based payments can be lowered to better approximate the marginal costs to the provider (Fujisawa and Lafortune 2008).

Indeed, there is a category of services that do not fit into the fixed-total versus activity-based dichotomy, called variously lump-sum payment or block grants. This approach has been common with hospital payment. In many OECD countries, hospitals receive block grants—contributions to budgets based on hospital size or type without specific regard to the number of or type of patients seen or services provided (Ellis and Miller 2008). In fact, unlike in the United States, in Europe, DRG payments often exist within a global budget set at the hospital level, with DRG payments representing 60 to 85 percent of revenues and block grants or other additional payments for certain high cost services making up most of the remainder (Quentin et al. 2010). The Medicaid program uses a similar approach to separating payment for services and lump-sum payments as supplements, as discussed in the discussion of lump-sum payments in the main text.

There have been examples of a similar approach for payment to clinicians, for example, fixed-amount payments for providers adopting electronic health records that meet "meaningful use" criteria. Some of the medical home demonstration models provided lump sum payments for practices to develop medical home infrastructure (Bitton, Martin, and Landon 2010).

Fixed payment approaches imply risk-bearing by the recipient organization – they are obligated to provide the services they have agreed to provide for the fixed payment amount. Some have portrayed the continuum of payment methods starting with no risk-bearing and moving to full risk-bearing evolving through intermediate steps, including bundled episodes, then one-sided shared savings, and then two-sided shared savings, and ending in full population-based payment, such as capitation. That has given rise to classification of payment systems based on a progressive assumption of risk in the payment models. The LAN payment model classification is the prototype of this approach that assumes a hierarchy of payment methods based basically on the progressive increase in magnitude of financial incentives (Health Care Payment Learning and Action Network 2017).
Appendix B. The U.S. Healthcare Primary Capitation Model

In the 1980s, HMOs pioneered new payment models for providers built on capitation. West Coast plans, in the shadow of Kaiser Permanente, adopted broad capitation to medical groups and newly formed independent practice associations (IPAs) which received PMPM payments for all professional services and in some cases for all health services (usually excluding prescription drugs). These were called professional capitation and global capitation, respectively. Under this approach in which the capitation payment would be made to a risk-bearing intermediary organization, the individual clinician usually was not paid capitation but rather salary in group and staff model HMOs, and fee schedules in network models.

The evolution of payment on the East Coast differed. U.S. Healthcare, a Pennsylvania-based HMO, that subsequently merged with Aetna, developed the prevailing model of primary care capitation that quickly spread among HMOs (Hanchak, Schlackman, and Harmon-Weiss 1996). In this model, network clinicians contracted directly with the HMO, rather than being part of an IPA. In this model, primary care physicians were paid on a capitated basis, with rates adjusted only for age and sex of members, who choose a physician to coordinate and regulate as a gatekeeper that member’s care. This model did not use health status-based risk adjustment.

Prior to 1987, U.S. Healthcare had a “withhold” model; each primary care physician received 80 percent of the capitated rate on a bimonthly basis. Based on analyzing the physician’s “risk pool”—the total cost of care provided during the course of the year, a portion of the total aggregate amount withheld across all primary care physicians would be returned at the year end in the form of a distribution. The proportion of the distribution for which each primary care practice was eligible depended on each physician’s performance in providing cost-effective care in the physician’s roster. The withhold amount provided a strong incentive to discourage overuse of services, especially to refer patients to specialists still paid on a fee service, rather than caring for them directly as part of their primary care responsibilities. In essence, the model was a particular strong form of shared savings based on performance of the risk pool, but without some of the technical approaches to limit the impact on a few, unavoidably high spending patients on the performance of the risk pool.

U.S. Healthcare itself modified the approach by adding a strong component of financial accountability for quality and basing shared savings less on the aggregate performance of all primary
care physicians and more directly on the individual physician's quality and total cost of care performance. The lead author of this paper contracted with an HMO using the U.S. Healthcare model, which also changed the model at about the same time. A key change was to eliminate the 20 percent withhold, responding to physician complaints. Instead, it reduced the capitation levels by 20 percent and offered bonuses based mostly on financial performance of the total cost of care risk pools. Technically, the payment model had eliminated financial risk, but had done so by reducing base payment. In modern parlance it moved from two-sided risk to upside-only but the effect on physician payment levels was the same. (Behavioral economics has found that penalties achieve more behavior change than bonuses. This model was implemented before behavioral economics.)

U.S. Healthcare and other HMOs developed fairly elaborate, quality-based approaches modifying the fairly crude initial model that was based only assessing financial performance on managing total cost of care. It moved to greater reliance on utilization metrics, as well as structure and process quality measures. The level of “risk” that was standard across HMOs far exceeded the levels that have been discussed today, such as with MIPs. The potential adverse incentives in pure capitation—that physicians would over-refer to non-risk bearing specialists—was recognized, while the approaches to try to reduce this behavior created other problems manifested by misleading total cost of care analyses because of the absence of risk adjustment for health status, and failure to adjust for the insurance risk of a small number of patients whose spending could not be influenced by involvement of primary care physicians.

Although the major backlash to managed care in the 1980s and 1990s did not seem to focus on capitation use in HMOs, at the time, there was concern that capitation, especially to an individual physician presents a conflict of interest for the physicians (Hanchak, Schlackman, and Harmon-Weiss 1996). In essence, as one of us wrote in 1991, HMOs took the same physicians who it thought were abusing fee-for-service incentives and trusted them not to underserve their capitated patients, while giving them direct incentives to do just that (Berenson 1991). The primary capitation models in use lacked a robust set of performance metrics that could be used to gauge a basic concern of capitation—stinting on care. Concern about stinting led some to recommend payment models then labeled “partial capitation” both to health plans and to individual clinicians so that organizations and physicians could act in their patients’ best interest, with little financial interest in patients’ choice of services (Newhouse 1994).
We consider a method to be a uniform approach to payment, whereas a model is a more complex package of one or more payment methods, with additional complementary components to influence its operational impact on clinician behavior. In this paper we first review common methods and then move on to a review of a number of payment models.


We accept the general, if imprecise, definition of value to mean more quality at lower cost.


Christopher Barbey, Nikhil Sahni, Robert Kocher, and Michael E. Chernew, "Physician Workforce Trends and Their Implications for Spending Growth."


The more precise US estimate, of 5.4 percent in 2016 by Phillips and colleagues, did not include spending for NPs, PAs and RNs. We used data from Phillips and colleagues based on the Medical Expenditure Panel Survey (MEPS), along with data from the US Bureau of Labor Statistics on the ratio of NPs to PAs (1.5) to distribute primary care spending for NPs and PAs. Using this, and the proportion of NPs and PAs that provide primary care that we presented earlier, we allocated primary care spending for NPs and PAs. We assume that the RNs represent a very small percentage of spending, mostly for level one office visits, so have excluded RNs from the calculations. Allocating NP and PA MEPS spending to primary care, we find that total primary care spending in the US is about 7 percent. However, MEPS calculations rely on household member recall of their health care encounters, producing a potential under-estimate for ambulatory care, including office visits, compared to hospitalizations and procedures, thereby underestimating primary care spending by an unknown amount. See Bernard DM et al., Reconciling Medical Expenditure Estimates from the MEPS and NHEA, 2012, available at https://meps.ahrq.gov/data_files/publications/workingpapers/wp_17003.pdf. See also "Occupational Employment and Wages, May 2017: Nurse Practitioners," US Bureau of Labor Statistics, accessed June 30, 2020, https://www.bls.gov/oes/2017/may/oes291171.htm; "Occupational Employment and Wages, May 2017: Physician Assistants," US Bureau of Labor Statistics, accessed June 30, 2020, https://www.bls.gov/oes/2017/may/oes291071.htm.

Some US definitions of primary care include OB/GYN, and others do not. We did not include OB/GYN in this estimate.
11 The OECD average of total doctors per 1,000 is 3.4. To calculate GPs per 1,000, we used data from the American Medical Association (AMA) Physician Masterfile (2017) to calculate the proportion of specialists (68 percent) and GPs (32 percent).

12 OECD has updated information for 2017, which we have calculated shows that the median ratio of specialty to general practitioner compensation for 21 countries is 1.29, with a fairly broad range (0.71–2.14). Unfortunately, the US was not included in this analysis.

13 We distinguish "bundled episode" payment from "episode" payment. The former refers to payment that covers all care over a defined time period for a clinical condition or procedure across all providers who provide care, whereas the latter refers to the duration of service the payment covers, whether or not provided by a single provider or by providers working together.

14 Per member per month (PMPM), which is the commonly used term, properly applies to those who have selected care in HMOs and does not apply to other insurance products where there is broad beneficiary choice of provider. Medicare beneficiaries are not members. Per person per month (PPPM) can be used generically or per beneficiary per month (PBPM) can be used specifically for Medicare beneficiaries. However, for consistency we will use PMPM, the common usage, in refer to all circumstances using capitation payment.

15 Physician net income is reasonably comparable to the work component of relative value units in the MPFS, which according to CMS in the past makes up about 50 percent of the total RVU value that is converted into the fee.

16 While commonly labeled “upside risk only,” the term is actually an oxymoron as by definition the approach does not include financial risk for losses. That is, there is no real financial risk in "upside only."

17 The Patient Protection and Affordable Care Act (ACA, P.L. 111-148, as amended) included DSH payment cuts, assuming that the expected increase in the number of people with health insurance (under both Medicaid and the health insurance exchanges) would lead to reductions in hospital uncompensated care, thereby lessening the need for DSH payments. The cuts have been delayed several times, although initially scheduled to take effect in FY 2014. See Medicaid and CHIP Payment and Access Commission March 2019 Report to Congress. https://www.macpac.gov/wp-content/uploads/2019/03/March-2019-Report-to-Congress-on-Medicaid-and-CHIP.pdf. See also "Disproportionate Share Hospital Payment," Medicaid and CHIP Payment and Access Commission, accessed May 25, 2020, https://www.macpac.gov/subtopic/disproportionate-share-hospital-payments/.


19 Federal Register vol. 84 no. 221, 62851–54 (Nov. 15, 2019).

20 Not all primary care physicians and other health professionals were eligible for the PCIP because the eligibility criteria were based on meeting a 60 percent threshold of allowed charges for specific, ambulatory E/M codes (excluding inpatient hospital and emergency department services). Some family physicians, particularly in rural areas, perform procedures, such that they did not meet the threshold to be considered primary care.


Physicians and physician offices, hospitals, critical access hospitals (CAHs), rural health clinics, federally qualified health centers, hospital-based or CAH-based renal dialysis centers, skilled nursing facilities, community mental health centers, renal dialysis facilities, homes of beneficiaries with end-stage renal disease getting home dialysis, and mobile stroke units.

Physicians, nurse practitioners, physician assistants, nurse-midwives, clinical nurse specialists, certified registered nurse anesthetists, clinical psychologists and clinical social workers, and registered dietitians or nutrition professionals.


Care originating from patient homes does not receive facility fees.

The national rates continue to be adjusted for geographic location and are reduced when there is a separate facility fee made.


"2018 National Health Center Data," Health Resources & Services Administration.


47 Readers are referred to https://www.annfammed.org/content/17/Suppl_1/S2 for a discussion of the implementation challenges many of these initiatives faced.

48 The six PCMH standards are enhancing access and continuity, identifying and managing patient populations, planning and managing care, providing self-care support and community resources, tracking and coordinating care, and measuring and improving performance.


50 This reflects available evaluations at the time of writing. For updated evaluations, see "Comprehensive Primary Care Plus," Centers for Medicare & Medicaid Services, https://innovation.cms.gov/innovation-models/comprehensive-primary-care-plus. See also "Comprehensive Primary Care Initiative," Centers for Medicare & Medicaid Services.

51 CMS and CMMI manage multiple types of ACOs, including the Medicare Shared Savings Program (MSSP), the ACO Investment Model for Medicare Shared Savings Program ACOs to test pre-paid savings in rural and underserved areas; the Advance Payment ACO Model for certain eligible providers already in or interested in the Medicare Shared Savings Program; the Comprehensive ESRD Care Initiative, for beneficiaries receiving dialysis services; the Next Generation ACO Model, for ACOs experienced in managing care for populations of patients; the Pioneer ACO Model (no longer active) for health care organizations and providers already experienced in coordinating care for patients across care settings; and the Vermont All-Payer ACO Model, to transform health care for Vermont's population.


The flat fee covers office/outpatient E&M services, prolonged E&M services, transitional care management, home care E&M, advanced care planning, Welcome to Medicare and Annual Wellness Visit, but did not address tele-health and e-visit services.


The two general primary care proposals are “An Innovative Model for Primary Care Office Payment,” submitted by Dr. Jean Antonucci and “Advanced Primary Care: A Foundational Alternative Payment Model (APC-APM) for Delivering Patient-Centered, Longitudinal, and Coordinated Care,” submitted by the American Academy of Family Physicians; "Advanced Care Model (ACM) Service Delivery and Advanced Alternative Payment Model", submitted by the Coalition to Transform Advanced Care, and the "Patient and Caregiver Support for Serious Illness", submitted by the American Academy of Hospice and Palliative Medicine. All proposals, PTAC Reports to the Secretary, and Secretary’s Response letters are available at https://aspe.hhs.gov/proposal-submissions-physician-focused-payment-model-technical-advisory-committee.


“International Health Care System Profiles,” Commonwealth Fund.

Anastasios Pappas, “Disease Management Programs (DMPs) in Germany: The Importance of Guideline Inclusion for Chronic Therapies,” Health Advances Blog, Health Advances, August 23, 2018,
This data does not differentiate capitation paid to IPAs and medical groups, which then distribute payment to physicians, often through fee schedules or salary, from primary care capitation payments made directly to physicians. Therefore, the percentage paid directly by primary care capitation is surely lower.


A recent issue brief (Gold, Green, and Westfall 2020) advocating a hybrid model similar to PCF asserted that a study (Basu et al. 2017) found that 63 percent of practice payment would need to be prospective to enable practice transformation. However, that interpretation is an overreading of the paper’s findings. The paper found the level of payments that would hold practices harmless financially under various assumptions of payment levels and a range of capitated payments before and after substitution of team and non-visit-based services for low complexity in-person visits. It is very specific to low volume primary care services and is focused on level of capitation needed to achieve neither financial gain nor loss for the practice, and not a more global assessment for achieving practice transformation.

In addition, lump sum payments typically pay for practice capacity development and are not related to actual services provided.

Note that this approach usually does not pay for time actually spent but rather based on the time committed to. Recent proposals for modifying fee schedule payments would pay for codes related to actual time spent.

References


Bindman, Andrew B., and Donald F. Cox. 2018. “Changes in Health Care Costs and Mortality Associated with Transitional Care Management Services after a Discharge among Medicare Beneficiaries.” *JAMA Internal Medicine* 178 (9): 1,283.


About the Authors

Robert Berenson joined Urban as an Institute fellow in 2003. He conducts research and provides policy analysis primarily on health care delivery issues, particularly related to Medicare payment policy, pricing power in commercial insurance markets, and new forms of health delivery based on reinvigorated primary care practices. In 2012, Berenson completed a three-year term on the Medicare Payment Advisory Commission, the last two years as vice chair. He also completed a term as an original appointee on the Physician-Focused Payment Model Technical Advisory Committee, established by the Medicare Access and CHIP Reauthorization Act of 2015. From 1998 to 2000, he was in charge of Medicare payment policy and private health plan contracting in the Centers for Medicare & Medicaid Services.

Adele Shartzer is a research associate in the Health Policy Center at the Urban Institute. Her work focuses on health coverage, access to care, and the health care delivery system, including value-based payment reform. She previously worked in the Office of the Assistant Secretary for Planning and Evaluation at the US Department of Health and Human Services. She received her PhD in health services research from the Johns Hopkins Bloomberg School of Public Health, where her dissertation research focused on the impact of hospital and insurer market concentration on consumers.

Roslyn C. Murray is a master’s student in public policy student at Georgetown University. Previously, she was the senior project and research manager at Catalyst for Payment Reform, a nonprofit organization with the mission to help purchasers and payers get better value for their health care dollars. A graduate of Stanford University, Murray holds a bachelor’s degree in human biology, with a concentration in health policy.
Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.