People dually enrolled in Medicare and Medicaid have, on average, more complex medical, long-term care, and social needs relative to people enrolled in only one of the programs. Because Medicare and Medicaid providers generally address different needs and lack incentives to coordinate with each other, dual enrollees risk missing needed services; receiving duplicative care; experiencing avoidable emergency department visits, hospitalizations, and readmissions; and exhibiting poor health outcomes. To reduce these risks, the Centers for Medicare & Medicaid Services (CMS), managed-care organizations, providers, and other stakeholders have developed various Medicare and Medicaid integrated care models. These models include the Program of All-Inclusive Care for the Elderly (PACE), Financial Alignment Initiative (FAI) models, Medicare Advantage Dual Eligible Special Needs Plans (D-SNPs), and Fully Integrated Dual Eligible Special Needs Plans.

Understanding the effects of integrated care plans on health outcomes, utilization, and spending remains an ongoing priority to improve care for dually enrolled people. CMS has funded formal evaluations of state FAI models, and other researchers have pursued various evaluative studies on other integrated care plans. In August 2020, the Medicaid and CHIP Payment and Access Commission (MACPAC) conducted a systematic literature review and created an inventory of 57 existing evaluations of integrated care models published between 2004 and July 2020. Their literature review indicated that most studies on integrated care plans found the plans decrease hospitalizations and
readmissions. However, findings related to other service types and outcomes are mixed and difficult to generalize across plan types, warranting continued research in this area.

What We Did

The purposes of this review are to critically assess the evaluative literature on integrated models for dual enrollees and to identify opportunities for researchers to consider. In particular, we examine the empirical and methodological approaches that have been used to evaluate integrated plans and highlight where gaps remain in the approaches taken. We reviewed each of the studies in the MACPAC inventory, focusing on the empirical methods used and the strengths and limitations of each approach.

What We Found

In the subsequent sections, we describe the findings from our review. First, we summarize the types of data sources and study designs commonly used. Then, we discuss the extent to which studies have incorporated a conceptual framework into their approaches and assess external validity. Though not the primary purpose of this review, we briefly summarize key findings from existing evaluations to help orient researchers seeking to contribute to the literature.

Data Sources Used

The most common data used in evaluation studies were Medicare and Medicaid administrative data, either from CMS, state Medicaid programs, or health plans. Commonly used CMS data include institutional and professional claims and encounters, as well as enrollment and demographic information, such as from the Medicare Master Beneficiary Summary base files, the Medicare-Medicaid Linked Enrollee Analytic Data Source, and Medicare Advantage monthly enrollment files. Several studies also used survey and assessment data from CMS, such as the Medicare Current Beneficiary Survey and the Long Term Care Minimum Data Set (standardized nursing home assessment data). Though highly detailed and comprehensive, these administrative data have drawbacks. They are unavailable for public use, expensive to access, and complex to use. Furthermore, they do not capture beneficiary experiences or other subjective but important measures of integrated plan success. Administrative data typically also do not include much demographic information beyond age, sex, and geography, or other pertinent information beyond that directly related to the data’s administrative function (e.g., income, wealth, household size, and marital status). Table 1 summarizes the key strengths and limitations of these data sources.

Five of the 57 studies reviewed relied entirely on qualitative data collected through focus groups, enrollee interviews, or enrollee surveys. Each of these studies focused on FAI models and aimed to describe enrollees’ experiences in and satisfaction with their integrated plans, as well as reasons for opting out. Though this type of data is necessary to paint a complete picture of how integrated care
plans affect enrollees, these data are also time consuming to collect and typically reflect a small sample of enrollees.

Other data sources used less often included Health Plan Management System payment files, medical records, nursing home admission records, all-payer claims datasets, the Area Health Resources Files, and complaints and appeals data.¹

**TABLE 1**

Key Strengths and Limitations of Commonly Used Data Sources for Studying Integrated Care Plans

<table>
<thead>
<tr>
<th>Data source</th>
<th>Key strengths</th>
<th>Key limitations</th>
<th>Studies that used data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS claims and enrollment data (e.g., Medicare and Medicaid claims)</td>
<td>Consistent, reliable administrative data on traditional Medicare and Medicaid enrollees, Contain service dates, diagnoses, procedures, and payments across all settings, Standardized across states, Extensive documentation available via the CCW</td>
<td>Only reflect diagnosed conditions, Contain limited clinical information, Contain limited demographic information, Expensive to obtain, Complex to use, Lag time (1–2+ years), Quality of Medicaid data varies across states</td>
<td>Beauchamp et al. (2008), Gattine, Jimenez, et al. (2019), Jones et al. (2013), Jung et al. (2015), Kane and Homyak (2004)</td>
</tr>
<tr>
<td>CMS assessment data (i.e., Long Term Care Minimum Data Set, home health Outcome and Assessment Information Set)</td>
<td>Describe enrollees in a specific setting (e.g., nursing facility, home health), Include functional and cognitive status data, Extensive documentation available via the CCW</td>
<td>Only represent enrollees when they reside in a specific setting, Some missing data problems, Expensive to obtain, Lag time (1–2+ years)</td>
<td>Bayer et al. (2018), Gattine, Elbaum-Williamson, et al. 2019, Ghosh et al. (2014), Holladay et al. 2018, Segelman et al. (2015)</td>
</tr>
<tr>
<td>Qualitative data (e.g., focus groups and interviews)</td>
<td>Reflect enrollee experiences and other elements unavailable via claims or assessment data</td>
<td>Usually unstructured, Smaller sample sizes, Time intensive to collect and analyze</td>
<td>Anderson et al. (2017), Craver, Cuellar, and Gimm (2016), Graham et al. (2018), McBride et al. (2017), Ptaszek et al. (2017)</td>
</tr>
</tbody>
</table>

Source: Authors’ review of select literature on integrated care plans for people dually enrolled in Medicare and Medicaid.  
Notes: CCW = Chronic Conditions Data Warehouse. Full citations appear in the reference list.

**Study Designs**

Research designs that randomly assign study participants to treatment versus control groups are the gold standard for estimating causal relationships. But in practice, people are rarely randomly assigned to participate or not participate in a program. Instead, a person's treatment or control status is often correlated with their future outcomes for reasons unrelated to their participation in or even eligibility for the program. This typically occurs for two reasons. First, if eligibility for the program is limited to a certain group of people (e.g., residents of one state), then the eligible population may differ from
ineligible populations in ways that correlate with future outcomes, which leads to "sample selection bias." Second, if participation among eligible people is voluntary, then those who choose to participate likely differ from those who choose not to participate in ways that correlate with their future outcomes, which leads to "self-selection bias." These two potential sources of bias are why researchers must rely on econometric and statistical techniques to estimate a causal effect in the absence of random assignment.

In the case of integrated care plans for dual enrollees, sample selection and self-selection bias must both be addressed by evaluators. Eligibility for integrated care plans is typically restricted to specific geographic areas and/or specific groups of dual enrollees. Furthermore, eligible people are typically allowed to choose whether they want to participate in an integrated care plan. In many cases, those eligible may opt into a new program; sometimes, enrollment into the integrated program is automatic, or "passive," but people can choose to opt out. In both cases (opting in or out), self-selection creates a significant issue.

Studies of integrated care plans acknowledged and addressed sample selection and/or self-selection to varying extents. We classified the studies from the MACPAC inventory into the following categories, based on the primary study design used: descriptive analyses, matching methods, difference-in-differences, and other. In the following paragraphs, we summarize the studies in these areas and discuss the threats to internal validity and the extent to which sample selection and/or self-selection bias were addressed in each study design. Table 2 summarizes the key strengths and limitations of these study designs.

**DESCRIPTIVE ANALYSES**

Several papers in the inventory described important aspects of integrated care plans, such as plan enrollment, plan characteristics, or enrollee experiences, but did not attempt to account for selection. Some studies did not incorporate a comparison group, or they incorporated a comparison group but did not adjust for differences between plan enrollees and nonenrollees. These descriptive studies relied on both qualitative and quantitative empirical methods. Qualitative methods commonly used include analyzing data collected from focus groups and interviews with individual enrollees. Other studies analyzed data from enrollee surveys and stakeholder interviews. For example, Graham, Liu, and Kaye (2016) compared telephone survey results among FAI demonstration enrollees with results for enrollees who dropped out, as well as dual enrollees living in nondemonstration counties. Examples of quantitative methods used in the descriptive literature include unadjusted calculations of enrollment rates by state and plan, pre-post analyses of costs and utilization, a comparison of utilization across five coverage models, and grade-of-membership methods (i.e., multivariate regression models to estimate relationships between sets of discrete variables).

**COMPARISONS USING MATCHING AND REGRESSION-BASED METHODS**

Matching is one method commonly used in studies seeking to measure the impacts of integrated plans for dual enrollees. The goal of matching is to reduce bias by identifying a comparison group of nonparticipants who are observably similar to participants and then comparing the outcomes between
the treatment group and the matched comparison group. However, matching methods do not account for self-selection on unobservable characteristics (i.e., factors that may drive a person's decision to enroll in or disenroll from an integrated care plan that are not observable in data and thus cannot be incorporated into the matching algorithm). Matching methods also do not eliminate sample selection bias if some unobservable characteristics are correlated with both the outcome studied and group status, and this is often a concern in empirical studies. This issue is more significant in studies that use administrative data sources, where demographic and other desirable information are limited. The internal validity of matching research designs thus hinges largely on how the comparison group is selected.

Nine studies created matched comparison groups of nonenrollees and used longitudinal analyses to compare outcomes over time between integrated care plan enrollees and the comparison group. In most of these studies, the matched comparison group was drawn from a group of people eligible to enroll in the integrated plan of interest and who had characteristics similar to enrollees but did not enroll. By comparing enrollees with people who were eligible to enroll but chose not to, estimates from these studies were susceptible to selection bias. Most such studies acknowledged this limitation. One study on the effects of PACE acknowledged that because enrollment in PACE requires people to give up their primary care physicians, differences in willingness to give up physicians (an unobservable trait) may drive enrollment decisions and be correlated with service-use habits. Kane and Homyak (2004) created two comparison groups: (1) people who lived in a FAI demonstration county but opted out and (2) people who lived in counties that were comparable with the demonstration counties but not covered by the demonstration. Creating a comparison group from people ineligible to enroll significantly mitigates this selection bias, and several matching studies took this approach. One study compared dual enrollees in a Medicare-Medicaid managed-care program in Massachusetts with a propensity-matched control cohort of dual enrollees in fee-for-service Medicare and Medicaid programs. Another study compared SCAN Health Plan enrollees with fee-for-service dual enrollees in California. Evaluations of PACE matched PACE enrollees to both (1) new enrollees in home- and community-based services 1915c waiver plans and (2) new nursing home entrants in the comparison group, because 1915c waiver participants had to meet standards for nursing home certifiability similar to PACE standards. However, ineligible people are likely less comparable with those who are eligible in other ways (e.g., if ineligible people live in a different geographic region or do not qualify for Medicare or Medicaid). This may be especially relevant for analyzing Medicaid service use, where the use of comparison groups in different states limits the ability to control for differences in the types of services available or the payment systems associated with other state Medicaid programs. However, even studies that used comparison groups drawn from areas within a demonstration state are often unable to control for within-state geographic variation in characteristics, such as provider supply or managed-care penetration.

Three additional studies compared enrollees with nonenrollees but adjusted for differences between the treatment and comparison groups using regression methods instead of matching
methods. The limitations outlined above generally apply to these regression-based methods. For example, an evaluation of the Minnesota Senior Health Options plan compared service use among enrollees with service use among Medicaid-only enrollees and Medicare-only enrollees, and it used a multivariate regression to control for observed characteristics of the two groups of enrollees.

**DIFFERENCE IN DIFFERENCES**

The evaluation studies in this review most commonly used a difference-in-differences (DID) research design. These research designs are advantageous for evaluating integrated care plans because they allow a person's participation status to be correlated with their baseline outcomes (e.g., spending, health status, or utilization). The key assumption for DID designs is that participation status is not otherwise correlated with changes in these outcomes over time. Such designs assume that in the absence of an integrated care plan, trends in outcomes between the treatment and comparison groups will evolve similarly over time. However, this assumption, known as the "parallel trends" assumption, is unlikely to be true in some cases (and is not directly testable). It is possible that people who opt in or out of integrated care plans differ from people who do not opt in or out in ways that are correlated with underlying trends in their outcomes over time (and are perhaps difficult to capture in data). Most DID evaluations acknowledged this possible limitation and attempted to mitigate its effects when the preintervention trends did not appear to be parallel.

Eighteen of the studies reviewed used DID designs. Fifteen of them are FAI demonstration evaluations prepared by the Research Triangle Institute, known as RTI, under contract for CMS that used an intent-to-treat DID approach. This approach compared the population eligible for a given demonstration with observably similar people meeting the same eligibility criteria but living where no demonstration model existed. Specifically, the treatment group included all dual enrollees eligible to enroll in a demonstration, even if they did not enroll or actively participate in the demonstration. The approaches used to create comparison groups varied slightly across state FAI evaluations. However, they generally involved selecting groups of counties with similar characteristics as demonstration counties (e.g., per beneficiary Medicare and Medicaid spending among dual enrollees, managed-care penetration, and patterns of institutional and community-based use of long-term services and supports) and then weighting people in those areas to approximate the demographic and other characteristics of people in demonstration areas. Including all eligible people in the treatment group, rather than only those who choose to participate, is intended to mitigate the effects of individual self-selection in these studies. That is, the treatment parameter in the intent-to-treat design estimated the average effect of having a FAI plan available among dual enrollees eligible to enroll in a FAI plan, regardless of actual participation in the plan.

Of the other three studies that used DID research designs, two included people who enrolled in the integrated care plan of interest in the treatment group and compared them with nonenrollees. Specifically, an assessment of Colorado's Accountable Care Collaborative included all collaborative members in the treatment group and all fee-for-service dual enrollees in the comparison group. Similarly, a study of the effects of a new coordinated care organization in Oregon included all managed-care organization enrollees who were automatically transitioned into coordinated care.
organizations in the treatment group and included fee-for-service enrollees in the comparison group. Finally, Zhang and Diana (2018) used a DID design with state-level D-SNP penetration as a continuous treatment variable. By relying on within-state variation in D-SNP penetration over time, this approach accounted for all baseline differences across states, but it also masked within-state heterogeneity.

OTHER STUDY DESIGNS

Several evaluation studies used study designs that do not fit into the above categories. Wieland and colleagues (2010) compared survival among PACE, waiver, and nursing home admissions using Kaplan Meier curves and risk stratification. Two studies prepared under contract for CMS used efficient orthogonal design—an experimental design approach—and regression analysis to test alternate ways of implementing components of care management. Lastly, savings reports prepared for CMS as part of the FAI evaluations compared trends in expenditures among FAI enrollees and a closed cohort of enrollees drawn from the comparison group used for the primary FAI evaluations.

TABLE 2

Key Strengths and Limitations of Commonly Used Study Designs for Studying Integrated Care Plans

<table>
<thead>
<tr>
<th>Study designs</th>
<th>Key strengths</th>
<th>Key limitations</th>
<th>Studies that used study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive analyses</td>
<td>Straightforward to implement and interpret</td>
<td>Cannot assess causality</td>
<td>Grabowski et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>May provide insight into enrollee experiences</td>
<td>Do not account for selection</td>
<td>Graham et al. (2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Graham, Liu, and Kaye (2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kim et al. (2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Schmitz et al. (2008)</td>
</tr>
<tr>
<td>Comparisons using matching and regression-based methods</td>
<td>Control for selection on observable characteristics</td>
<td>Do not control for selection on unobservable characteristics</td>
<td>Foster et al. (2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Jones et al. (2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kane and Homyak (2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Meret-Hanke (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rawal and Munear (2012)</td>
</tr>
<tr>
<td>Difference in differences</td>
<td>Control for selection on observable and unobservable characteristics at baseline</td>
<td>Biased if parallel trends assumption does not hold</td>
<td>Gattine et al. (2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Require panel data</td>
<td>Gattine, Jimenez, et al. (2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Justice et al. (2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ormond et al. (2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Zhang and Diana (2018)</td>
</tr>
</tbody>
</table>

Source: Authors’ review of select literature on integrated care plans for people dually enrolled in Medicare and Medicaid.

Note: Full citations appear in the reference list.

Incorporation of Conceptual Frameworks

Few studies of integrated care models included coherent theories of change or conceptual justifications for why and how one should expect the integrated care plan to affect beneficiary outcomes. Rarely did studies include explicit, testable hypotheses. Instead, most studies described or evaluated the effects of the plan overall without carefully considering why the plans might have had the effects they did or which plan elements were most influential.
Two studies prepared under contract for the US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, tested the effects of specific mechanisms within integrated care plans. These studies used an experimental research design, where participating care teams were assigned to implement different combinations of routine or enhanced care across more than 10 intervention components; outcomes were compared among beneficiaries who received routine versus enhanced care in each component. The tested components reflected very specific elements of care coordination, and the outcomes were designed to be logically related to the interventions. For example, one of the components was “fall prevention referral,” where the intervention involved sending a letter to members in addition to routine fall prevention care, and the targeted outcome was emergency department visits and readmissions.  

One other study of PACE explicitly listed and tested hypotheses for how each of its independent variables (individual-level risk factors and program-level variables) would affect enrollee health outcomes. One independent variable was a dichotomous indicator for whether the program had a medical director trained in geriatrics, and the associated hypothesis was that a medical director who had received this specialized training might provide superior care. The study found several program characteristics to be associated with prevention of functional decline but few to be associated with health and mortality.

EXTERNAL VALIDITY
To assess the external validity of studies, we considered several dimensions of generalizability: geographic coverage, study population, and study time frame.

Most evaluations focused on integrated care programs within one state. Notably, the 15 FAI evaluations focused on state-specific FAI models. In some cases, the models were only available to select counties or geographic areas within a state.

Few studies examined integrated care programs across state lines. The PACE evaluation reports prepared for CMS examined the effects of PACE in eight states with sufficient numbers of PACE enrollees. Only one study (Zhang and Diana 2018) examined effects across all states and Washington, DC. However, because Medicaid programs vary across states, we do not expect that results obtained in one state will necessarily be replicable in another state. This variability in regulatory environments underscores the importance of understanding how individual components of a model work and thus increase the value of an explicit conceptual framework.

The eligible population for integrated care plans in the evaluation studies varied. For FAI demonstrations, some states included all full-benefit dual enrollees over age 18 or 21, whereas others restricted the eligible population to those ages 21 to 64 or ages 65 and older. Some states, such as Washington State, targeted their demonstrations to high-cost, high-risk enrollees. These variations in study population limit the external validity of FAI evaluation results. Evaluations of PACE plans were, by definition, limited to community-dwelling elderly dual enrollees with medical frailty. Thus, findings from PACE evaluations may not be generalizable to other types of dual enrollees served by integrated care models.
The time frame of existing evaluation studies also varied. The 15 FAI evaluations were all released in or after 2017, including several in the past year. On the other hand, several PACE evaluations used data from more than 10 years ago. Zhang and Diana’s study of early D-SNPs also used nearly 10-year-old data, whereas other studies used data from the mid-2010s. Though the health problems and social needs of dual enrollees as a group may not change dramatically over time, the nature of the markets for medical care and long-term services and supports do change as new models of care arise, new payment models are created (e.g., the merit-based incentive payment system), and other regulations change. Thus, estimates from a decade ago may no longer be relevant.

Study Findings
The evaluation studies identified in this review focused primarily on the effects of integrated care models on spending and utilization outcomes available from claims and encounter data. Some studies also examined program enrollment, enrollee experiences, quality of care, and health outcomes. In general, findings across all domains tended to be mixed, suggesting underlying heterogeneity in the design of integrated models, the populations targeted or enrolled, or the providers delivering care under the models.

SPENDING
The estimated effects of integrated care programs on spending were mixed. Several studies found integrated care plans reduce Medicare spending. However, some found no significant changes to spending or had mixed findings. The mixed findings even among the FAI managed-care demonstrations, for which care models and evaluations had largely uniform designs, suggest that either the Medicare cost savings generated by integrated care are not large, or differences in area characteristics, contextual factors, and implementation across individual managed-care plans may drive results. To the extent these differences have not been accounted for in evaluation designs, studies with additional controls may be appropriate.

To date, published studies have largely been unable to estimate the effects of integrated care on Medicaid spending because of a lack of high-quality data. Two earlier studies of PACE that used only data from the Medicaid Statistical Information System, or MSIS, found mixed results on Medicaid spending, but the difference in methods between the studies makes comparison difficult. A key promise of Transformed Medicaid Statistical Information System, or T-MSIS, data has been a timelier and more accurate picture of Medicaid spending and utilization. These data are new, however, and the extant literature has not yet included them in analyses. For the purposes of FAI demonstration evaluations, the DID design will require a mix of data from the MSIS and T-MSIS in every demonstration, and the comparability of these data systems has yet to be studied.

UTILIZATION
Findings about service utilization varied across studies. Numerous studies found reductions in inpatient stays, readmissions, emergency department visits, and long-stay nursing facility use or use of long-term services and supports. However, other studies found few changes in the use of
some or all of these services, and some studies even found increases in inpatient stays and readmissions, emergency department use, and long-stay nursing facility use. The inconsistencies across study findings here—particularly the findings of opposite-signed and statistically significant effects on readmissions and emergency department and nursing facility use—also suggest the idiosyncrasies of individual demonstrations and plans require more study.

ENROLLMENT
Evaluation studies that examined enrollment found enrollment and participation to be lower than expected, even in programs with passive enrollment. For example, an analysis of demonstration enrollment across eight states found that enrollment rates ranged from 5 to 62 percent as of October 2016, despite passive enrollment into demonstrations in these states. Several descriptive studies aimed to understand people’s enrollment decisions surrounding integrated care plans. These studies found that the complexity or lack of information and concerns about continuity of care (e.g., concerns about changing doctors) were key reasons people chose not to enroll. Administrative problems, such as difficulty contacting enrollees, also contributed to low enrollment.

ENROLLEE EXPERIENCES
Many of the findings in the descriptive literature pertained to enrollees’ experiences with integrated care programs. Several studies documented that enrollees were generally satisfied with their integrated care plan or the care coordination they received through it. Some studies found that those who used care coordination services reported increased access to care; however, several studies also reported enrollee concerns about access to specialty services (e.g., behavioral health), durable medical equipment, prescriptions, and/or other long-term services and supports.

Studies that examined quality of care and health outcomes generally found little to no improvement in these outcomes. Findings of satisfaction among integrated care participants without strong evidence of improvements in quality or health outcomes, along with low rates of voluntary participation, suggest several possibilities: The benefits of integrated care may not be well communicated to potential enrollees. Alternatively, unobserved characteristics of self-selected enrollees may make them more likely to both enroll in and be happy with the new care models. On the other hand, the standard measures of quality and health outcomes used in these studies may not be those that determine satisfaction, and other measures should be studied. However, many quality and outcome metrics require electronic medical records or survey data linked to enrollment data for integrated models, neither of which are generally available for large-scale studies. It may be feasible to conduct smaller-scale studies where such data are available.
Conclusions and Suggestions for Researchers

Integrated care plans are designed to improve care, avoid unnecessary hospitalizations, and reduce spending for people dually enrolled in Medicare and Medicaid. For more than a decade, policymakers and researchers have sought to understand the degree to which integrated care plans have achieved these goals. The complexity and heterogeneity of the plans and their implementation create a challenging framework for evaluation, and more research is needed to understand these plans’ effects.

In this literature assessment, we summarized the common data sources used and methodological approaches taken in existing evaluation studies of integrated care plans. We found that many evaluations relied on matching or DID research designs, which are widely used in the health economics literature but have limited abilities to control for enrollee self-selection into integrated care plans. We also found that few studies included conceptual frameworks or tested specific hypotheses related to the plans, making it difficult to pinpoint the specific element or elements of plans that were effective or not. Finally, many evaluation studies were limited to plans in only one state or are now more than 10 years old, limiting the generalizability of their findings to other regions or to today’s health care environment.

We suggest that researchers looking to advance the evaluative literature on integrated care plans consider creative methodological approaches to addressing selection bias. Researchers may consider using instrumental variables or regression discontinuity design approaches, which have not been used to date to evaluate integrated care plans but may be promising approaches to reduce the effects of selection bias. Future studies would also benefit from clearly defining the hypotheses being tested, such that the proposed causal pathway is clearly delineated and findings can be attributed to a specific element of integrated care plans.

The needs for increased care coordination and integration of Medicare and Medicaid services remain important priorities for improving the health of dual enrollees and reducing unnecessary health care utilization and spending. Despite a growing number of evaluative studies on integrated care plans finding evidence that the plans effectively reduce hospitalizations, the somewhat mixed findings related to other outcomes and the heterogeneity in the designs of the plans suggest more research is needed. By incorporating comprehensive data, employing methodological approaches that minimize bias to estimate causal effects, explicitly testing hypotheses, and carefully considering the generalizability of findings, additional research can further elucidate the aspects of integrated care plans that are effective or not in achieving their goals.
Notes

1 See MACPAC (2019).

2 See Anderson, Feng, and Long (2016); Bayer et al. (2018, 2019); Beauchamp et al. (2008); Foster et al. (2007); Gattine et al. (2016); Gattine, Elbaum-Williamson, et al. (2019); Gattine, Jimenez, et al. (2019); Ghosh, Schmitz, and Brown (2014); Holladay et al. (2018); JEN Associates Inc. (2013); Jones et al. (2013); Jung et al. (2015); Justice et al. (2016, 2018); Kane and Homyak (2004); Meret-Hanke (2011); Rawal and Munevar (2012); Schmitz et al. (2008); Segelman et al. (2015); Walsh et al. (2016, 2017, 2018); Walsh, Sandler, Zhao, Simms, et al. (2019); Walsh, Sandler, Zhao, Trapnell, et al. (2019); Wilkin et al. (2017); Zhang and Diana (2018); and Zurovac et al. (2014a, 2014b).

3 See Bayer et al. (2018); Gattine, Elbaum-Williamson, et al. (2019); Ghosh, Schmitz, and Brown (2014); Holladay et al. (2018); JEN Associates Inc. (2013); Justice et al. (2018); Segelman et al. (2015); and Walsh et al. (2018).

4 See Anderson et al. (2017); Craver, Cuellar, and Gimm (2016); Graham et al. (2018); McBride et al. (2017); and Ptaszek et al. (2017).

5 See Grabowski et al. (2017); Kim et al. (2019); Segelman (2014); Vouri et al. (2015); Walsh et al. (2016); Wenger et al. (2011); and Wieland et al. (2010, 2013).

6 See Anderson et al. (2017); Graham et al. (2018); McBride et al. (2017); Ptaszek et al. (2017); and Schmitz et al. (2008).

7 See Graham et al. (2018) and Kim et al. (2019).

8 See Anderson et al. (2017); Craver, Cuellar, and Gimm (2016); and Graham, Liu, and Kaye (2016).


10 See Grabowski et al. (2017).


12 See Kim et al. (2019).

13 See Wieland et al. (2013).

14 See Beauchamp et al. (2008); Foster et al. (2007); Ghosh, Schmitz, and Brown (2014); JEN Associates Inc. (2013); Jones et al. (2013); Meret-Hanke (2011); Kane and Homyak (2004); Kane et al. (2006); and Rawal and Munevar (2012).

15 See Foster et al. (2007).

16 See Kane and Homyak (2004).


18 See Rawal and Munevar (2012).

19 See Beauchamp et al. (2008) and Ghosh, Schmitz, and Brown (2014).

20 See Anderson, Feng, and Long (2016); Jung et al. (2015); and Segelman et al. (2015).


22 See Bayer et al. (2018, 2019); Gattine et al. (2016); Gattine, Jimenez, et al. (2019); Griffin et al. (2019); Holladay et al. (2018, 2019); Justice et al. (2016, 2018, 2019); Ormond et al. (2018, 2019); Sandler et al. (2019); and Walsh et al. (2018).

23 See Lindrooth et al. (2016).

24 See Kim and Charlesworth (2016).

26 See Wieland et al. (2010).
27 See Zurovac et al. (2014a, 2014b).
28 See Sandler et al. (2018); Walsh et al. (2017); Walsh, Sandler, Zhao, Trapnell, et al. (2019); and Wilkin et al. (2017).
29 See Kim and Charlesworth (2016) and Walsh, Sandler, Zhao, Simms, et al. (2019).
30 See Zurovac et al. (2014a).
31 See Anderson, Feng, and Long (2016); Bayer et al. (2018, 2019); Gattine et al. (2016); Gattine, Elbaum-Williamson, et al. (2019); Gattine, Jimenez, et al. (2019); Griffin et al. (2019); Holladay et al. (2018, 2019); JEN Associates Inc. (2013); Jones et al. (2013); Jung et al. (2015); Justice et al. (2016, 2018, 2019); Kane et al. (2016); Kane and Homyak (2004); Kim and Charlesworth (2016); Lindrooth et al. (2016); Ormond et al. (2018, 2019); Rawal and Munevar (2012); Sandler et al. (2018); Walsh et al. (2017, 2018); Walsh, Sandler, Zhao, Trapnell, et al. (2019); and Wilkin et al. (2017).
32 See Bayer et al. (2018) and Holladay et al. (2018).
33 See Beauchamp et al. (2008); Foster et al. (2007); and Ghosh, Schmitz, and Brown (2014).
34 See Zhang and Diana (2018).
35 See Bayer et al. (2018).
36 See Bayer et al. (2019); Holladay et al. (2018); and Ormond et al. (2018).
38 See Walsh et al. (2018).
39 See Justice et al. (2019).
40 See Griffin et al. (2019); Holladay et al. (2019); Justice et al. (2019); Ormond et al. (2019); and Walsh, Sandler, Zhao, Simms, et al. (2019).
41 See Beauchamp et al. (2008); Foster et al. (2007); Kane et al. (2016); Kane and Homyak (2004); Meret-Hanke (2011); and Segelman et al. (2015).
42 See Zhang and Diana (2018).
43 See Kim and Charlesworth (2016) and Lindrooth et al. (2016).
44 See Bayer et al. (2019); Holladay et al. (2018); Justice et al. (2018, 2019); Lindrooth et al. (2016); Sandler et al. (2018); Wilkin et al. (2017); and Zhang and Diana (2018).
45 See Bayer et al. (2018); Gattine, Jimenez, et al. (2019); Holladay et al. (2019); and Ormond et al. (2019).
46 See Foster et al. (2007); Ghosh, Schmitz, and Brown (2014); Kane and Homyak (2004); Walsh et al. (2017); and Walsh, Sandler, Zhao, Trapnell et al. (2019).
47 See Foster et al. (2007) and Wieland et al. (2013).
48 See Anderson, Feng, and Long (2016); Bayer et al. (2018); Beauchamp et al. (2008); Holladay et al. (2018, 2019); Justice et al. (2018); Meret-Hanke (2011); Ormond et al. (2019); and Rawal and Munevar (2012).
49 See Anderson, Feng, and Long (2016); Holladay et al. (2018) and (2019); and Walsh et al. (2018).
50 See Anderson, Feng, and Long (2016); Bayer et al. (2018); Gattine, Jimenez, et al. (2019); JEN Associates Inc. (2013); Justice et al. (2018, 2019); Kane and Homyak (2004); Lindrooth et al. (2016); and Segelman et al. (2015).
51 See Jung et al. (2015); Justice et al. (2018); Kane et al. (2016); and Kane and Homyak (2004).
52 See Gattine, Jimenez, et al. (2019) and Justice et al. (2019).
See Bayer et al. (2018) and Justice et al. (2019).

See Holladay et al. (2018) and Ormond et al. (2019).

See Bayer et al. (2019); Grabowski et al. (2017); Graham et al. (2018); Graham, Liu, and Kaye (2016); Holladay et al. (2018); and Ormond et al. (2018).

See Grabowski et al. (2017).

See Grabowski et al. (2017); Graham et al. (2018); Graham, Liu, and Kaye (2016); and McBride et al. (2017).

See Grabowski et al. (2017).

See Anderson et al. (2017); Bayer et al. (2019); Craver, Cuellar, and Gimm (2016); Graham et al. (2018); Graham, Liu, and Kaye (2016); Holladay et al. (2019); Justice et al. (2019); Ormond et al. (2018, 2019); and Ptaszek et al. (2017).

See Ginsburg and Eng (2009); Graham et al. (2018); and Kim et al. (2019).

See Anderson et al. (2017); Graham et al. (2018); McBride et al. (2017); and Ptaszek et al. (2017).

Bayer et al. (2019); Beauchamp et al. (2008); Ghosh, Schmitz, and Brown (2014); Kane et al. (2016); Kim and Charlesworth (2016); and Wieland et al. (2010).

References


Kim, Hyunjee, and Christina Charlesworth. 2016. *Assessing the Effects of a Coordinated Care Organization on Dual-Eligibles in Oregon*. Portland, OR: Oregon Health and Science University, Center for Health Systems Effectiveness.

Lindrooth, Richard, Gregory Tung, Tatiane Santos, Rose Hardy, and Sean O'Leary. 2016. Evaluation of the Accountable Care Collaborative. Aurora, CO: University of Colorado Anschutz Medical Campus.


About the Authors

Laura Barrie Smith is a research associate in the Health Policy Center at the Urban Institute. Her research examines the effects of health care policy interventions using large claims and electronic health record datasets paired with quasi-experimental research designs. She studies topics related to access to care, provider behavior, the primary care workforce, health IT, and health equity. Smith holds a BA in mathematics from St. Olaf College and a PhD in Health Services Research from the University of Minnesota.

Timothy A. Waidmann is a senior fellow in the Health Policy Center. He has over 20 years of experience designing and conducting studies on varied health policy topics, including disability and health among the elderly; Medicare and Medicaid policy; disability and employment; public health and prevention; health status and access to health care in vulnerable populations; health care utilization among high-cost, high-risk populations; geographic variation in health care needs and utilization; and the relationships between health and a wide variety of economic and social factors. Waidmann’s publications based on these studies have appeared in high-profile academic and policy journals. He has also been involved in several large-scale federal evaluation studies of health system reforms, assuming a central role in the design and execution of the quantitative analyses for those evaluations. Before joining Urban in 1996, Waidmann was assistant professor in the School of Public Health and postdoctoral fellow in the Survey Research Center at the University of Michigan. He received his PhD in economics from the University of Michigan in 1991.

Kyle J. Caswell is a senior research associate in the Health Policy Center. His research covers multiple areas related to health and economic well-being, with a focus on vulnerable populations. He is currently working with colleagues to evaluate a demonstration to coordinate health care for dually eligible Medicare-Medicaid beneficiaries, and on a study to evaluate how disability status affects Medicare spending among the elderly. Previous projects include an evaluation of economic well-being among elderly individuals with mental health impairments and disability insurance, the financial burden of medical spending, the impact of managed care among Medicaid beneficiaries, uncompensated health care, and inequalities in health outcomes. Before joining Urban, Caswell was an economist in the US Census Bureau’s Health and Disability Statistics Branch, where he contributed to the medical out-of-pocket spending component of the Supplemental Poverty Measure. During his previous tenure at Urban, he worked with colleagues to develop estimates of potential savings in medical spending attributable to preventive health services. Caswell holds a PhD in economics.
Acknowledgments

This work was supported by Arnold Ventures. 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.

We also thank DaQuan Lawrence and Clair O’Brien for their assistance in assembling and organizing the literature for this review.