MELBOURNE INSTITUTE
Applied Economic & Social Research

Working Paper Series

The impact of value-based payment reform on medical expenditures, fees and volume of services. Early evidence from a large-scale fee schedule reform in Australia

Dajung Jun
Anthony Scott

Working Paper No. 4/22
February 2022

THE UNIVERSITY OF MELBOURNE
MELBOURNE INSTITUTE
APPLIED ECONOMIC & SOCIAL RESEARCH
60 YEARS IMPACT

The impact of value-based payment reform on medical expenditures, fees and volume of services. Early evidence from a large-scale fee schedule reform in Australia*

Dajung Jun
Melbourne Institute: Applied Economic & Social Research,
The University of Melbourne

Anthony Scott
Melbourne Institute: Applied Economic & Social Research,
The University of Melbourne

Melbourne Institute Working Paper No. 4/22
February 2022

*The research was funded by the NHMRC Partnership Centre on Health System Sustainability (9100002). We would like to thank all seminar participants at the Australian Health Economics Society and the International Health Economics Association in 2021, and feedback from Jongsay Yong.

Melbourne Institute: Applied Economic & Social Research
The University of Melbourne
Victoria 3010 Australia
Telephone **+61 3 8344 2100**
Fax **+61 3 8344 2111**
Email **melb-inst@unimelb.edu.au**

Website **melbourneinstitute.unimelb.edu.au**

Abstract

Value-based payment reform helps ensure that payments from governments and insurers provide incentives to support the provision of high value healthcare. This research evaluates the early effects of the first major review of the Australian Medicare Benefits Schedule (MBS), the MBS Review, that subsidises medical services provided by general practitioners (GPs) and non-GP specialists in private practice. A key objective of the MBS Review was to increase the provision of value-based health care. We evaluate the overall impact of changes implemented in the first three and a half years of the review (2016 to 2019) on total medical expenditures, volume of services, and fees. We use data from 2,216 doctors who responded to the Medicine in Australia: Balancing Employment and Life (MABEL) panel survey and consented to have their MABEL data linked to MBS claims between 2011 and 2019. Between 2016 and June 2019, 8.5% of all unique MBS items had been changed by the Department of Health. Of the 488 items affected, 183 items (37.5%) were removed, 122 (25%) were added, and 183 were amended (37.5%). We use recently developed difference-in-difference methods accounting for staggered adoption and heterogeneous treatment effects. We find no statistically significant effects on medical expenditures, the volume of care, or average fees charged. The exception is for GPs, where we found evidence of a 10% increase in average fees that can be sourced to an increase in the fee for spirometry diagnosis in 2018 to encourage the provision of more appropriate care for respiratory diseases. The scale of the MBS Review was substantial, and the net overall effect was cost neutral. Further research needs to examine the longer-term effects of the MBS Review.

JEL classification: I10, I13, I18

Keywords: Physicians, fee-for-service, healthcare costs, health insurance

1. Introduction

Governments and insurers who fund health care need to decide what services are covered and subsidized and how to use such funding to support high-value healthcare using value-based payment models. These decisions influence access to healthcare for the population, health outcomes, and health expenditures through their effects on patient and provider behaviours. In fee-for-service settings, fee schedules of health insurers are used to fund medical services and these fee schedules usually evolve incrementally over time as new technologies that attract government subsidies are added. Previous research from the United States has examined changes to Medicare fees where these fees are fixed by Medicare. With fixed fees, changes to fees have been shown to lead to changes in the volume of services provided as providers seek to maintain or increase their incomes (Clemens & Gottlieb, 2014; Rice, 1983). Larger scale changes to fee schedules have also occurred. The most notable is the introduction of the Resource-Based Relative Value Scale (RBRVS) in U.S. Medicare that changed the relative levels of fees to reflect the relative costs of providing services (Hsiao et al., 1992).

More recent changes to provider payment models have been undertaken alongside an increased awareness of the need to improve quality of care through reducing the provision of low value care and increasing the under-provision of high-value care. Though the term 'value' is used in the health policy literature it has been largely interpreted by the medical profession in terms of the effectiveness of treatment rather than its cost-effectiveness. For example, low value care has been interpreted and measured as care of negligible effectiveness and that could cause harm and so is 'unnecessary'. High value care is care that is effective but could be under provided. Low value care includes overdiagnosis and overtreatment, as well as "pricing waste," with estimates of around 30 % of health care spending as waste (Berwick & Hackbarth, 2012; Shrank, Rogstad, & Parekh, 2019). Governments and insurers in some countries have been attempting to re-align their payment models in the light of evidence around quality and the effectiveness of care (Scott, Liu, & Yong, 2018; Song et al., 2012). The reform of provider payment schedules is one way to improve value and efficiency in health care.

This paper aims to examine the early effects of a major review of the national fee schedule in Australia. Medicare is Australia's national universal tax-financed insurance scheme for the whole population. A main part of Medicare includes subsidies for privately-provided medical services through the Medicare Benefits Schedule (MBS). Physicians are paid by fee-for-service and are allowed to set their own fees. The MBS specifies the level of government subsidy for each medical service (item) provided and subsidizes part of patients' out-of-pocket costs rather than directly reimbursing doctors.

In 2015 the MBS Review Taskforce was established to undertake the first major review of the MBS since Medicare was introduced in 1984. The aims of the MBS Review were wide ranging but included a focus on reducing low value care, providing more appropriate care and modernising the MBS to better support improved health outcomes. Seventy percent of items on the MBS had not been changed since their introduction, and most of these were not subjected to an evidence-based review that is required for all new items before they are added to the MBS. *"An important objective of the Review will be curbing inefficiency by ensuring that low-value services—that is, services which provide no or negligible clinical benefit and, in some cases, might actually do harm to patients—cease being funded…with the overarching goal of promoting the provision of the best patient outcomes for our health expenditure*" (MBS Review Taskforce, 2015). The terms of reference include the following: *"Review MBS items taking account of factors including concerns about safety, clinically unnecessary service provision and accepted clinical guidelines"*. Though reducing low value care should reduce costs on a treatment-by-treatment basis, the MBS Review also included the adoption of evidence-based clinical guidelines that might increase the provision of effective healthcare in the population. There was no specific mandate to reduce costs

At the start of the review in 2015, the MBS included 5,769 items. The Taskforce intended to review all items and was led by clinicians, with the review delivering its final report in December 2020 (MBS Review Taskforce, 2020). As of August 2021, the government has agreed to 817 taskforce recommendations and implemented more than 580 recommendations, involving changes to over 3,000 MBS items (Department of Health, 2021). Implementation of recommendations is expected to continue until 2023.

This aim of the paper is to examine the early impact (up to July 2019) of the MBS Review on health expenditures for private medical services provided by General Practitioners (GPs) and non-GP specialists[1]. We also examine some of the mechanisms of the effects, such as changes in the volume of services provided and changes in doctor's fees. Unlike Medicare in the United

---

[1] In Australia GPs are considered to be specialists.

States, a particular feature in Australia is that the MBS provides subsidies for patients, while doctors are free to charge above the MBS fee and can price discriminate. Doctors can potentially offset any effects on their revenue from changes to the MBS by changing their volume, mix of billings or fees charged. The MBS Review Taskforce was led by clinicians and involved the removal of items, the addition of new items, the amendment of item descriptions, changes to the rules about who and in what circumstances items could be claimed, changes to fees, and the bundling/unbundling of items. Since the taskforce was led by clinicians, it might be unlikely that they would recommend changes that reduce their own revenue. Our hypothesis is that health expenditures for private medical services will remain the same or increase.

We have rich data on doctors and their characteristics from a unique panel survey of doctors linked to MBS claims, and so are also able to examine heterogeneity in the effects of the reform to examine if certain types of a doctor were more likely to respond than others. We use unique data that matched the survey responses of 2,000 doctors who completed the Medicine in Australia: Balancing Employment and Life (MABEL) panel survey of doctors to doctors' MBS claims between January 2012 and July 2019. During this period, recommendations of the MBS Review were implemented at roughly quarterly intervals from 2016. We use a difference-in-difference (DID) study design to examine the overall impact of the reform using recently developed methods by (Callaway & Sant'Anna, 2021) for staggered adoption designs.

2. Institutional Setting

Australia spends 10% of its GDP on healthcare and has a rich mix of public and private finance and provision (AIHW, 2020). Medicare is a tax-funded universal health insurance scheme providing healthcare access to all Australians, including free access to public hospitals and subsidized access to pharmaceuticals and private medical services. The MBS is a fee schedule used to subsidize private medical services provided by General Practitioners (GPs) and non-GP specialists (e.g., obstetricians, surgeons, paediatricians, psychiatrists) in private practice. There is no enrolment or registration of patients with GPs, but patients require a referral from a GP to see another specialist. The fee charged by a doctor can be higher than the Medicare schedule fee, which is fixed for each item and does not vary by geography. The Medicare benefit (or rebate) is the subsidy paid to patients for each item. The Medicare benefit is equal to 100% of the Medicare schedule fee for consultations provided by GPs, is 85% of the Medicare schedule fee for all other out-of-hospital services, such as consultations provided by non-GP specialists, and is 75% of the Medicare schedule fee for in-hospital services provided to private patients. Doctors are free to charge above the Medicare schedule fee, in which case patients face an out-of-pocket cost. Doctors can also choose a bulk bill where their fee is equal to the MBS benefit paid, so patients do not pay any out-of-pocket cost. Around 85% of services provided by GPs are bulk billed, compared to approximately 30% of services offered by other specialists.

For out-of-hospital services, mainly GP and specialist consultations, the difference between the MBS benefit paid for each service and the doctor's fee is the patient's out-of-pocket cost. This cannot be covered by private health insurance. Still, a Medicare Safety Net provides additional subsidies for those who incur out-of-pocket costs above certain thresholds in a calendar year, but only for out-of-hospital services. For private in-hospital services, the gap between the Medicare schedule fee and the MBS benefit (75% of the Medicare schedule fee) is covered by private health insurance, and the amount above the Medicare schedule fee is paid by the patient and/or the private health insurer depending on the policy.[2]

The MBS was historically based on fee schedules that evolved during the 1970s and was expanded in 1984 when Medicare was introduced. There was a review of the MBS in Australia based on the RBRVS reforms that occurred in U.S. Medicare, but it was not implemented (National Centre for Classification in Health, 2000; Wright, 2015). MBS fees for existing items have been increased by inflation most of the time. The Medical Services Advisory Committee (MSAC) usually decides on changes to the MBS using an evidence-based approach, mainly the addition of new items reflecting new technologies. MBS fees for services are recommended by MSAC but negotiated between the Department of Health and the medical profession. In 2015 a major review of the MBS was announced, with the establishment of the MBS Review Taskforce that would recommend changes to the MBS independently of MSAC. Over 100 clinical committees were established to review groups of items by clinical discipline (MBS Review Taskforce, 2020). A Principles and Rules Committee was also established that made general recommendations about the appropriate use of different items (e.g., some items cannot

---

[2] Private insurance can be used to cover part or all the remaining gap but only if the doctor has an agreement with the patients' private health insurer. This agreement can be for 'no gap' cover where there is no out of pocket payment to the patient or 'known gap cover' where there is an agreed out of pocket payment. Similar to bulk billing, it is at the doctor's discretion as to whether to use this 'gap cover' agreement. Patients have zero out-of-pocket medical costs (i.e., 'no gap' cover is used, or the service is bulk billed) for around 50% of private in-hospital services. The federal government also subsidises private health insurance, with around 45% of the population holding cover, and which can be held and used in parallel with Medicare.

be claimed together or in succession) and a Consumers Committee. For recommendations agreed by the government, Implementation Liaison Groups were established to inform implementation and further consider any unintended consequences. The final change to any item could therefore be different to what was originally recommended by the Taskforce.

The emphasis was on the use of evidence which could be used to reduce over-utilised low value care or increase the provision of under-utilised high value care. (Elshaug, Watt, Mundy, & Willis, 2012). 'Value-based' health care has generally been defined in purely medical terms that focussed on whether care provided was effective. There was no specific mandate to reduce costs, and there was no explicit mandate to use cost-effectiveness analysis when making recommendations. The terms of reference did not mention low value care though this was mentioned in the initial consultation paper as a key aim of the review. The terms of reference, talked about 'reducing clinically unnecessary care', and providing advice on 'appropriateness', and 'best practice options'. In practice, this meant that the scope was wider than the removal of low value care items such that it was not necessarily the case that costs would fall. The type of changes that occurred to items included the removal of obsolete items, the addition of new items sometimes to replace old ones, bundling/unbundling of items, changes to MBS fees, and changes to the circumstances (rules) under which doctors or patients can claim items or in which settings. The clinical committees made over 1,400 recommendations from 60 reports to the Taskforce, which then reviewed these and made recommendations to the government. In August 2021, 817 recommendations had been agreed to be implemented, and a further 539 are being considered. Implementation has occurred at irregular time intervals, roughly quarterly since January 2016 and will continue until 2023.

3. Conceptual framework

The aim of the paper is to examine the impact of the implementation of MBS Review Taskforce recommendations on health expenditure for private medical services, comprising the numbers of each item claimed multiplied by the fee charged by the doctor for each item summed over all items. This total includes expenditures by Medicare, patients, and private health insurers. There are various types of changes to each item, each of which has potentially different effects on expenditure. Clearly the removal of items representing low value care would reduce spending but many other changes to existing items were also made to increase the provision of under-utilised effective medical care such that the net effect on total spending is ambiguous. Sometimes several different types of changes affect a single item (e.g., a change to the description of services intended to be provided and an increase in the MBS fee). In contrast, a single change can influence a group of items (e.g. preventing doctors from claiming a consultation item at the same time as a procedural item – co-claiming). Since clinicians led the review and made all recommendations, we assume that they did so in a way that would not reduce their revenue. In addition, the Department of Health who approved the recommendations and decided which ones to implement would also be conscious of not increasing overall costs. Our main null hypothesis is that the net effect of all changes on medical expenditures is zero.

How doctors respond to these changes will also be heterogeneous. Doctors are assumed to optimize their billing practices, the volumes of services provided, and their fees to maximize their revenue or at least maintain it where changes to the MBS are made. Behavioural responses of doctors to changes to specific items will depend on the marginal costs of changing their clinical and/or billing behaviour relative to the marginal benefits in terms of expected changes to revenue and effects of patients health status. Being able to carefully optimise billing to maintain revenue can be costly and is more likely for doctors who work in larger medical groups that use medical billing companies to manage billing and have more administrative support, and less likely for doctors who work on their own or who have smaller private practices and manage their own billing. The mechanisms through which expenditure might change are complex because of the heterogeneity of the types of changes across different specialties and the high degree of product differentiation, as well as dynamics due to the differential timing of changes. Each specific change will have a different behavioural impact on the total expenditure on each affected item, which is specific to each item, through effects on the volume of claims for that item or effects on the fee charged per item, including decisions on whether to bulk bill.

There could also be heterogeneity across different types of doctors. For example, those who are more altruistic and care less about maximizing profits will be less likely to change their behaviour when items are changed. Overseas trained doctors comprise a significant part of Australia's health care system, as they are an effective alternative to address the lack of a medical workforce in rural communities or remote areas. Any differences between them could be attributable to a lack of knowledge about the Australian health care system and Medicare billing, cultural and communication issues, or the differences in medical training style from overseas. Pre-existing literature also shows that gender gaps exist and females are often paid less, which could be the result of well-documented differences in practice styles, hours worked, and productivity (Gravelle, Hole, & Santos, 2011; Roter, Hall, & Aoki, 2002; Schurer, Kuehnle, Scott, & Cheng, 2016; Valle et al., 2019). If one item is changed, doctors could respond by altering the volume and/or fees of that item but also of other unrelated items claimed by the same or different patients. Such spillovers are more likely for changes to items that have larger effects on revenue. These potentially unintended consequences are difficult to predict a priori.

4. Data

As our main focus is to examine the overall impact of the MBS Review Taskforce recommendations, we first manually create the data that includes all implemented changes from January 2016 to June 2019. Each deleted, introduced, or changed item by the MBS Review Taskforce is identified. More details of how we create this part of the data are in the Appendix.

The implemented changes are summarized in Table 1. Between 2016 and June 2019, 8.5% (488/5,769) of all unique MBS items (5,769 as at baseline in 2015) had been changed by the Department of Health (Table 1). Of the 488 items affected, 183 items (37.5%) were removed, 122 (25%) were added, and 183 were amended (37.5%). Among the amended items, item descriptions were changed for 61.2% of these changes, 20.8% had their fees changed, and 18% had the circumstances when an item could be claimed altered (e.g., added restrictions or block co-claiming).[3] For all changes that happened within our timeframe, there are five implementation periods, and the number of unique items affected at each period varied from 7 to 283.

This item-level dataset of changes recommended by the MBS Review Taskforce and implemented by the Department of Health is linked to MBS claims data from 2,216 respondents of the Medicine in Australia: Balancing Employment and Life (MABEL) panel survey. MABEL collects data from approximately 10,000 doctors per year from 2008 to 2018 (Joyce et al., 2010; Szawlowski, Harrap, Leahy, & Scott, 2020). This includes rich data on doctor characteristics. In 2015, we requested consent from around 12,000 qualified GPs and non-GP specialists who had previously completed a MABEL survey to link their MABEL survey

[3] Co-claiming refers to the situation when a procedure is performed and the doctor also claims other items at the same time (often consultation items). MBS Review Taskforce recommended disallowing co-claiming for some surgical items.

responses to all of their MBS claims: 2,216 (around half GPs and half other specialists) gave permission for linkage. These represent 4.2% (2,216/52,299) of all GPs and non-GP specialists in Australia in 2015 undertaking clinical practice[4]. In 2015, their linked MBS data included 6,384,829 items, representing 1.76% of all MBS items claimed in that year.

Our estimation sample included the changes identified in Table 1 that appeared in our data. First, we include three out of five implementation periods: November 2017, March 2018, and November 2018. None of the changes in July 2016 appeared in our sample and were, in any case, mainly the removal of 'obsolete' items no longer used and so with very low volumes of claims. Although we have a few observations in our data affected by July 2018 changes, these were not sufficient for analysis and so were excluded. Furthermore, other items identified in Table 1 may not appear in our dataset. For example, in the three periods, Table 1 identifies 159 items that were deleted. In our estimation sample, only two items are identified as deleted because our sample of doctors did not claim 157 of the deleted items before they were deleted.

The MBS claims data are at the item level. The information includes the date when the service was provided, doctors fee (provider charge), MBS fee (schedule fee), and MBS benefit paid, and a unique doctor identifier that is linked to the MABEL survey. For each doctor, we have their complete claims history across all the locations in which they practice. Since this does not include patient characteristics or a patient identifier, we cannot identify which items were claimed by the same patient. These data included around 3.48 million observations for these survey respondents between November 2011 to June 2019. For our analysis, we aggregate each item to the level of item-month-doctor. This provides us with 2.42 million observations.

5. Empirical approach

We first recognize that multiple interventions occurred at each time point and across multiple time points and we need to account for this staggered adoption and heterogenous treatment effects. Which items were changed depend on various factors: (1) when each clinical committee was set up, (2) the process used to make its recommendations to the MBS Review Taskforce, (3) the process by which the Taskforce made recommendations to the Department of Health, and (4) the process used by the Department of Health and the Minister to decide which

---

[4] From our sample frame provided by AMPCo.

recommendations to implement. The timing of the changes was subject to similar internal bureaucratic processes. The Implementation Liason Groups play a role that influences the timing of actual implementation. Some randomness in the bureaucratic processes could exist across items, but it could be that the initial selection of items or decisions to implement some recommendations were based on the volume and expenditure of the affected items as well as evidence around low-value care (Elshaug et al., 2012). Clinical areas or items with health expenditures or high volumes could have been indirectly used to set these priorities by clinical committees, the Taskforce, or the Department of Health. This would be based on the data and other evidence presented to each committee on the volume of services and expected effects of recommended changes. The selection of items to change and the variation in implementation timing could therefore be non-random and potentially correlated with the dependent variables.

In our case, the MBS Review was implemented at three different time points in our data, and the analysis needs to account for this staggered adoption. Each treatment effect at each time point will be heterogeneous since a different set of items, and potentially a different set of doctors are affected at each time of implementation. In addition, each treatment effect would be estimated using different time periods in the data, such that for later treatment effects, we observe a shorter time period of data after the change than earlier treatment effects.

Since the interventions are characterized by staggered adoption and heterogeneous treatment effects, we use a newly developed DiD estimator that overcomes known biases from using the standard two-way fixed effects (TWFE) approach (Callaway & Sant'Anna, 2021). The standard TWFE model estimates a single average treatment effect on the treated (ATT). Goodman-Bacon (2021) has shown how the ATT, which is a weighted average of each "group time" ATT, will be biased and have no clear causal interpretation if each treatment effect changes over time, potentially resulting in negative weights leading to the wrong sign on the DiD regression coefficient. Goodman-Bacon (2021) demonstrates these weights depend on two factors: (1) differences in the number of treated and untreated units at each time point and (2) differences in the variance of the treatment effects due to differences in the length of time for each treatment. If there are two treatment groups (an early treated group and a late treated group), it is unclear how to interpret the ATT from a TWFE analysis since a range of possible comparisons are being made: i) a comparison of the early treated group to the units that are never treated, ii) a comparison of the late treated group to units that are never treated, iii) a comparison of the early treated group to the late treated group before it was treated, and iv) a comparison of the late treated group to the early treated group after the treatment.

A range of new methods applicable in different settings have been developed and are reviewed by De Chaisemartin and D'Haultfoeuille (2021) and Roth, Sant'Anna, Bilinskiz, and Poe (2022). We use the new approach developed by Callaway and Sant'Anna (2021), who have developed a semi-parametric DiD estimator for staggered adoption designs with multiple periods, variations in treatment timing, and heterogeneous treatment effects. This method corrects the issues with negative weights and heterogenous treatment effects in TWFE models by defining a 'group-time average treatment effect' $ATT(g,t)$ where *g* refers to the group of units (where a unit in our data is each affected Medicare item per doctor per month) affected by the intervention at time *t* (months). Each $ATT(g,t)$ is estimated separately and can be easily aggregated in various ways to define the policy-relevant treatment effects of interest. Callaway and Sant'Anna (2021) assume that if the unit is treated, then it is treated afterward without being "turned off' again. In our analysis, we assume no anticipatory effects. The $ATT(g,t)$ is defined as the following:

$$ATT(g,t) = E[Y_t(g) - Y_t(0)|G_g] = 1, for\ t > g \tag{1}$$

This definition of the ATT is for units (MBS items) that are members of a group (*g*) at a particular time period (*t*). G defines which group the units belong to with G_g a binary variable that is equal to one if a unit is first treated. $Y_t(0)$ denote each unit's untreated potential outcome at time *t* if they remain untreated through to period T; in other words, if they were not to participate in the treatment across all available time periods. $Y_t(g)$ denotes the potential outcome that each unit would experience at time *t* if they were first treated in group *g*. This ATT varies across calendar time (in our case, each month) and group. The conditional parallel trend assumption is below.

For each $g < t$,

$$E\left[Y_t(0) - Y_{t-1}(0)\middle|X, G_g = 1\right] = E[Y_t(0) - Y_{t-1}(0)|X, C = 1] \tag{2}$$

where *C* is defined as a binary variable that is equal to one for units that do not participate in the treatment in any time period (the never treated). This assumption holds after conditioning on *X*, the observable characteristics of each unit. Defining C as never treated units avoids the issue of early-treated units becoming controls for later-treated units, and where in the last period, not all units are treated.[5] Callaway and Sant'Anna (2021) propose estimation procedures based on outcome regression, inverse probability weighting (IPW), or doubly robust methods that use both. We use doubly robust methods as they require only one of either the outcome regression or the IPW method to be correctly specified. This method also assumes that the generalized propensity score is uniformly bounded away from 1. In our context, the MBS item is the treated or untreated unit, and the time period is each month.[6]

The observable characteristics of each item need to be defined as pre-treatment covariates and so be time-invariant. These include specialty (GP or specialist), doctor's gender, whether one's date of birth is earlier than 1960, the place they work (major city vs. rural area), and whether the doctor is qualified overseas. We also include each item's share of the total volume of claims for each doctor in 2015, as well as whether the time is classified in three broad item groups defined in the MBS (Consultation items, Diagnostic Tests and Pathology, and Obstetrics, Anaesthetics, Operations, others). One thing to note regarding those covariates is we always have a non-zero number of affected items by the MBS Review across all characteristics that are used as covariates (see Table 2), and so this validates the common support assumption that the covariate distributions overlap between the treated and untreated units.

Callaway and Sant'Anna (2021) propose a straightforward way to aggregate each group-time average treatment effect to understand heterogeneity in the effect of participating in the treatment. To consider heterogeneous effects across groups, we consider the following parameter (using their notation):

$$\theta_{sel}(\tilde{g}) = \frac{1}{T - \tilde{g} + 1} \sum_{t=\tilde{g}}^{T} \mathrm{ATT}(\tilde{g}, t) \qquad (3)$$

5 Another option is using `not-yet treated' group as the control group. We test this as a robustness check.

6 We cannot have the doctor as the treated unit since at each time point other changes are made to the MBS (e.g., recommendations for new items from MSAC) in addition to item changes associated with the MBS Review.

$\theta_{sel}(\tilde{g})$ is the ATT among units in group $\tilde{g}$, after we sum across all their post-treatment periods. This gives the 'single' group-time ATT for each group across all of their post-treatment periods. In our case, we are primarily interested in the total ATT across the whole time period, and so the single ATTs, $\theta_{sel}(\tilde{g})$, can simply be added together across all groups:

$$\theta_{sel}^{O} = \sum_{g \in G} \theta_{sel}(g) w_{sel}^{O}(g,\ t) \tag{4}$$

θ_{sel}^{O} is the ATT for those who have ever been treated from equation (3) and $w_{sel}^{O}(g,\ t)$ is a weight, $P(G = g | G \leq \mathrm{T})$, that accounts for the length of time that each unit is treated. As we are interested in the overall effect of the MBS Review, this estimate would be a meaningful way to summarize the group effects of the MBS Review. Callaway and Sant'Anna (2021) recommend a bootstrap procedure that accounts for multiple hypothesis testing and clustering for asymptotically valid inference.

6. Results

Table A1 in the Appendix compares the doctors in our sample to that of the population of doctors in clinical practice at the time they consented in 2015. There are similar proportions of GPs and specialists, but doctors in our sample are more likely to be female, older, from outside a major city, and to have qualified in Australia. These variables are used as control variables in the analysis. It is more appropriate to examine representativeness in terms of the distribution of the number of items claimed in our data since items are the unit of analysis. Tables A2, A3 and A4 compare the number and distribution of items claimed for our sample compared to national claims data (separately for GPs, non-GP specialists, and claims for diagnostic imaging and pathology) in 2015 when the doctors provided consent and before items were changed by the MBS Review. Within each of these groups, the distribution of claims is broadly representative of all items claimed in 2015.[7]

[7] Where there are major differences, these are for claims made by certain types of doctor who are much less likely to complete the MABEL survey. For GPs (Table A1) our sample underrepresents claims for services provided after hours. This is because at that time many of these claims were made by GPs working for commercial deputising services, and these GPs were less likely to complete the MABEL survey. For non-GP specialists (Table A3) the main difference is for item group 'A2 Other non-referred attendances'. These are for 'Other Medical Practitioners' who are doctors who have not achieved Fellowship of a college but are allowed to claim

Table 2 shows descriptive statistics in the matched dataset from the pooled estimation sample, including the survey responses from doctors, claims they made, and the implemented changes by the MBS Review Taskforce. Panel A shows the number and proportion of claims, pooled over all years, across doctor characteristics.[8] For example, for all items claimed per month, on average 53.6% were claimed by GPs. Panel B presents the item-level characteristics. Except for the "share of the item for each doctor in 2015," which measures the proportion of each doctors' total claims accounted for by each item, variables in Panel B are the outcomes used in our analysis. Our main focus is the total medical expenditure for each item per month per doctor, defined from our data as the sum of the doctors' fees charged for each item for each month. We also use total expenditure by Medicare, which is the sum of Medicare benefits paid per item per doctor per month. To investigate the potential mechanisms, we examine effects on the average doctors' fee (price) per item per doctor per month, the number of claims (volume) per item per doctor per month, and the proportion of claims for each item that are bulk billed per doctor per month.

In Panels C1 and C2, we divide the sample based on whether each item was affected by the MBS Review changes at any time of the analysis. In Panel C1, the share of affected items per doctor in 2015 was 15.5 %, and the total medical expenditure is $2,860 per item per doctor per month. The total number of claims per item in Panel C1 is 24.5. Most of the mean values (except the proportion of the bulk-billed claims) in Panel C2 are lower than those in Panel C1, suggesting that the affected items had higher expenditures, volumes, and average fees than items not affected. Among changes identified in our merged dataset, 86.4% were amendments, while 9.7% and 3.9 % were additions or deletions, respectively. Of the different changes implemented, a higher proportion of claims involved adding items than deleting them, whilst the majority of changes were amendments to existing items.

Table 3 examines the descriptive statistics for the changes by type of change (i.e., amendment, deletion, and addition) one-year before and after implementation (a simple event study type comparison). For example, if item 102 was changed in July 2018, we included data 12 months before and after July 2018. If item 103 was changed in November 2018, we included data 12

MBS items. These are usually GPs in training or overseas GPs allowed to work in rural areas or commercial deputising services. These doctors are less likely to complete MABEL.

[8] We chose 1960 since this is the median age of our sample of doctors across the period. Since the initial timing of each doctor's entry to our MABEL sample varies, instead of using age (which is time-variant), we w use the time-invariant characteristic the year of birth.

months before and after November 2018. Then we averaged these across all items that had changed. For those items that changed, we observed slight increases after implementing amendments in most measures except the proportion bulk-billed. For deleted items, we only have two in March 2018 in our final dataset: items 32090 and 32093. These deletions were made to address substantial national variation in the use of colonoscopy, and the expenditures and average fee for these items were relatively higher than the average shown in Panel C1 of Table 2. In contrast, the newly added items in the last panel of Table 3 show lower volume and medical expenditures than those which were amended or deleted. This may reflect slow adoption of these new items.

Table 4 examines the characteristics of the items that were affected, as well as the characteristics of doctors who were affected. Whilst consultation items are the most common item claimed, 10.9% of diagnostic testing and pathology items were affected by the MBS Review, followed by 7% of obstetrics, anaesthetics, operations, and other items, and 5.4% of consultation items. Only 3% of claims made by GPs were affected by the MBS Review Taskforce recommendations, compared to 11.6% of items claimed by other specialists during our analytic timeframe. Other than the doctor's specialty, there are negligible differences in affected items by other doctor characteristics. Male, younger doctors, doctors in major cities, and those from overseas were slightly more likely to be affected.

Table 5 shows the overall ATT and the group-specific effects. The results show a 3.8% fall in total medical expenditures, a 4.5% fall in Medicare expenditure, and 3.4% fall in the number of claims, and much smaller changes to bulk-billing rates and average fees. Though these are reasonably sized effects, none are statistically significant. The separate group effects at each of the three-time points also show no statistical significance.

Table 6 examines the heterogeneity of the results depending on the doctors' characteristics. In Panel A, we examine the differential effects between GPs and specialists. We do not find significant differences in most of the impacts of the Taskforce recommendations across doctor characteristics. The only statistically significant increase that we observe is in average fees among GPs, which increased by 10%. On further investigation, this was attributable to a single change that affected GPs in November 2018 for spirometry (lung function testing for patients with asthma, chronic obstructive pulmonary disease, or other lung diseases). A new item (11505) was introduced for annual spirometry testing to confirm diagnosis with a schedule fee of $41.10, which was intended to replace an old item (11506, schedule fee $20.55) but which was retained and changed to be used for monitoring of obstructive airway diseases (Department of Health, 2019). This, therefore, represented an effective doubling of the fee as GPs switched to the new item for this very common test conducted in primary care. The intention was to encourage more appropriate care and monitoring of patients with lung disease. Note that though the average fee was higher, medical expenditure for GPs did not increase overall, and bulk-billing of items did not change.

Table 7 presents the robustness of the overall ATT results. First, we replace our control group from the never-treated to the not-yet-treated. We then use either IPW or outcome regression. The results are similar. When we do not include any time-invariant covariates and essentially just compare unconditional mean differences, the coefficients are slightly larger and statistically significant. Still, the treated and untreated units will have different characteristics that will be unaccounted for. In rows (5) and (6), we examine whether excluding some observations with a very high number of claims per item-doctor-month would change our results. In our dataset, four doctors (one dermatologist and three pathologists) made 3,000 claims per month for particular items related to diagnostic tests, which can be of high volume but could skew the results.[9] Row (5) excludes all claims made by those four doctors, and the results are similar. In row (6), to address the concerns about some pathologists and radiologists who might routinely have a very high volume of claims compared to other doctors, we exclude all doctors and claims from these specialties, and the results remain similar. Since we have a relatively long pre-treatment period (from October 2011 to June 2016), we examine whether a shorter pre-period changes our results. Row (7) limits the time period from August 2014, which is two years before the first implementation (July 2016, see Table 1).

7. Discussion

Many richer countries are using payment methods to support the provision of high value healthcare. This is the first study providing evidence of a major large-scale review of the national fee schedule for doctors in Australia. We examine evidence of overall impact on medical expenditures, volume and fees for the early part of the MBS Review, from 2016 to June 2019. Across the various dimensions and outcomes, we consistently observe no statistically significant effects on medical care expenditures, though effect sizes suggest

[9] The item categories that are related with this significant number of claims are “P1 Haematology” , “P2 Chemical”, “P3 Microbiology” and “P10 Patient Episode Initiation.”

reductions in medical spending of around 4% which is reasonably large. The exception is for GPs where one specific change to spirometry led to higher average fees overall.

There are various reasons for these overall null effects. As we examined the overall effects, there may have been impacts from specific changes, but that these were offset by other changes, such that the net effect is close to zero. It may also be the case that the changes implemented first where those that caused least disruption to billing and revenues or were of low volume, and more substantial or controversial changes have yet to be introduced, have been delayed because of protracted negotiations, have not been implemented at all, or have been introduced after July 2019 when our data ends. In addition, some changes to some items were 'piloted' first with an intention to review again at a later stage.

Since clinicians drove the recommendations, it might be expected that they did not make recommendations that reduced their revenue from Medicare. Even if they made recommendations that increased their earnings, the process of implementation by the Department of Health and stakeholder groups might have reduced the likelihood of significant potential increases in medical expenditures. In addition, since doctors can alter the volumes of services provided as well as fees, they can optimise their billing behaviours such that their earnings are at least maintained. This is different from U.S. Medicare with fixed prices, where doctors control only the volume of care, and so any changes to fees can only be offset by changes to the volume.

A question remains as to whether the MBS Review improved the provision of value-based health care. It is positive that the MBS Review did not lead to increases in medical care expenditures. We have no data on health outcomes or the provision of low-value health care. The changes implemented were motivated by a desire to reduce low-value care, but many other factors would have shaped the recommendations and outcomes of the review. The majority (86%) of the changes made were amendments to existing items, largely changing descriptions of items. Changing how items are described in the MBS has previously been shown to have no impact on behaviours (Scott & Shiell, 1997) as they provide no financial incentives to alter behaviour, and the relatively small proportion of claims that involved deletions or additions of items may have cancelled each other out in terms of changes to overall expenditures.

Our dataset is too small a sample of doctors to pick up large changes within specific specialties, and so further research needs to use larger datasets to enable us to examine the impacts of single item changes within specific specialties where there may have been effects. Though our sample is not representative of doctors, it is broadly representative of items claimed in the pre-reform year of 2015. Items claimed are the unit of analysis in our estimation sample, and we show that doctor characteristics play little role in which items are affected by the MBS Review (Table 4) and in influencing treatment effects (Table 6). This suggests that doctors were affected equally by the review no matter their gender, age, rurality, or whether overseas qualified, suggesting that differences in the characteristics with the population of doctors are unlikely to affect the generalisability of our results.

The overarching goal of value-based payment reform across countries is to provide more necessary care without significant cost increases. To achieve these aims, countries such as the U.S. have been experimenting with and implementing payment reform for decades with often mixed results on costs and quality of care (Scott et al., 2018). Our results provide the first evidence that costs may not have increased as a result of the MBS Review. We contribute to the literature by employing relatively new econometric methods for staggered adoption designs. As this major reform was implemented sequentially instead of happening at once, the new econometric method that we adopt in this paper enabled us to identify its effect more accurately than traditional methods. This study is also the first to examine the early effects of the MBS Review, heralded as representing a major change to Australian Medicare. Further research is required to examine the longer-term impacts.

References

AIHW. (2020). *Health expenditure Australia 2018-19* Canberra: Australian Institute of Health and Welfare.

Berwick, D. M., & Hackbarth, A. D. (2012). Eliminating waste in US health care. *Journal of the American Medical Association, 307*(14), 1513-1516.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics, in press*. doi:10.1016/j.jeconom.2020.12.001

Clemens, J., & Gottlieb, J. D. (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *The American Economic Review, 104*(4), 1320.

De Chaisemartin, C., & D'Haultfoeuille, X. (2021). Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. *SSRN Electronic Journal*. doi:10.2139/ssrn.3980758

Department of Health. (2019). Changes to diagnostic services for respiratory function tests. Retrieved from http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Factsheet-RespiratoryFunctionTests

Department of Health. (2021). *The provision of general practitioner and related primary health services to outer metropolitan, rural and regional Australians. Submission from the Department of Health to the Senate Standing Committee on Community Affairs*. Canberra: Australian Government.

Elshaug, A. G., Watt, A. M., Mundy, L., & Willis, C. D. (2012). Over 150 potentially low-value health care practices: an Australian study. *Med J Aust, 197*(10), 556-560.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. In press. *Journal of Econometrics*. doi:10.1016/j.jeconom.2021.03.014

Gravelle, H., Hole, A. R., & Santos, R. (2011). Measuring and testing for gender discrimination in physician pay: English family doctors. *Journal of Health Economics, 30*(4), 660-674. doi:10.1016/j.jhealeco.2011.05.005

Hsiao, W. C., Braun, P., Dunn, D. L., Becker, E. R., Yntema, D., Verrilli, D. K., . . . Chen, S. P. (1992). An overview of the development and refinement of the Resource-Based Relative Value Scale. The foundation for reform of U.S. physician payment. *Medical Care, 30*(11 Suppl), NS1-12. doi:10.1097/00005650-199211001-00001

Joyce, C. M., Scott, A., Jeon, S.-H., Humphreys, J., Kalb, G., Witt, J., & Leahy, A. (2010). The "Medicine in Australia: Balancing Employment and Life (MABEL)" longitudinal survey - Protocol and baseline data for a prospective cohort study of Australian doctors' workforce participation. *BMC Health Services Research, 10*(1), 50. doi:10.1186/1472-6963-10-50

MBS Review Taskforce. (2015). *Public Submissions Consultation Paper*. Canberra: Australian Government Department of Health.

MBS Review Taskforce. (2020). *An MBS for the 21st Century Recommendations, Learnings and Ideas for the Future. December*. Canberra: Australian Government Department of Health.

National Centre for Classification in Health. (2000). *Professional Relativities Study Final Report*. Sydney: National Centre for Classification in Health.

Rice, T. H. (1983). The Impact of Changing Medicare Reimbursement Rates on Physician-Induced Demand. *Medical Care, 21*(8), 803-815. doi:10.2307/3764772

Roter, D. L., Hall, J. A., & Aoki, Y. (2002). Physician Gender Effects in Medical Communication. *JAMA, 288*(6), 756. doi:10.1001/jama.288.6.756

Roth, J., Sant'Anna, P., Bilinskiz, A., & Poe, J. (2022). *What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature*.

Schurer, S., Kuehnle, D., Scott, A., & Cheng, T. C. (2016). A Man's Blessing or a Woman's Curse? The Family Earnings Gap of Doctors. *Industrial Relations: A Journal of Economy and Society, 55*(3), 385-414.

Scott, A., Liu, M., & Yong, J. (2018). Financial Incentives to Encourage Value-Based Health Care. *Medical Care Research and Review, 75*(1), 3-32. doi:10.1177/1077558716676594

Scott, A., & Shiell, A. (1997). Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics, 16*(3), 323-342. doi:10.1016/S0167-6296(96)00520-6

Shrank, W. H., Rogstad, T. L., & Parekh, N. (2019). Waste in the US Health Care System: Estimated Costs and Potential for Savings. *JAMA, 322*(15), 1501-1509. doi:10.1001/jama.2019.13978

Song, Z., Safran, D. G., Landon, B. E., Landrum, M. B., He, Y., Mechanic, R. E., . . . Chernew, M. E. (2012). The ‘Alternative Quality Contract,’ Based On A Global Budget, Lowered Medical Spending And Improved Quality. *Health Affairs, 31*(8), 1885-1894. doi:10.1377/hlthaff.2012.0327

Szawlowski, S., Harrap, B., Leahy, A., & Scott, A. (2020). *Medicine in Australia: Balancing Employment and Life (MABEL). MABEL User Manual: Wave 11 Release*. Melbourne: Melbourne Institute: Applied Economic and Social Research, The University of Melbourne.

Valle, L., Weng, J., Jagsi, R., Chu, F.-I., Ahmad, S., Steinberg, M., & Raldow, A. (2019). Assessment of Differences in Clinical Activity and Medicare Payments Among Female and Male Radiation Oncologists. *JAMA Network Open, 2*(3), e190932. doi:10.1001/jamanetworkopen.2019.0932

Wright, M. (2015). Is there value in the Relative Value Study? Caution before Australian Medicare reform. *Med J Aust, 203*(8), 331-333. doi:10.5694/mja15.00571

Table 1. Changes based on the MBS Review Taskforce recommendations

Change date	Number of items affected	Item categories	Type of changes (number of each type of change)
July 2016	26	Diagnostic Radiology (I3); Miscellaneous Diagnostic Procedures (D1); Nuclear Medicine Imaging (I4); Obstetrics (T4); Regional or Field Nerve Blocks (T7); Surgical Operations (T8); Miscellaneous Therapeutic Procedures (T1)	Deletion (24); Descriptions Amended (2)
November 2017	142	Surgical operations (T8); Assistance at operations (T9); Obstetrics (T4); Miscellaneous Diagnostic Procedures and Investigations (D1); Consultant Physician Attendances To Which No Other Item Applies (A4); Specialist Attendances To Which No Other Item Applies (A3); Relative Value Guide For Anaesthesia (T10); Pain and Palliative Medicine (A24); Consultant Occupational Physician Attendance (A12); Sexual Health Medicine (A32); Addiction Medicine (A31); Diagnostic Radiology (I3)	Addition (11); Deletion (42); Descriptions Amended (26); Fee Amended (34); Restriction changes or block claiming (29)
March 2018	30	Urgent Attendance After Hours (A11); Surgical operations (T8)	Addition (24); Deletion (4); Descriptions Amended (2)
July 2018	7	Surgical Operations (T8)	Addition (3); Descriptions Amended (4)
November 2018	283	Computed Tomography (I2); Diagnostic Radiology (I3); Miscellaneous Diagnostic Procedures and Investigations (D1); Magnetic Resonance Imaging (I5); Regional or Field Nerve Blocks (O11); Miscellaneous Therapeutic Procedures (T1); Relative Value Guide for Anaesthesia - Medicare Benefits are Only Payable for Anaesthesia Performed in Association with an Eligible Service (T10); Surgical Operations (T8)	Addition (84); Deletion (113); Descriptions Amended (78); Fees Amended (4); Block claiming (4)

Notes: The group number in parentheses in the third column are item categories as assigned by the MBS. The number in parentheses under the last column is the corresponding number of changes. This includes the information up until 2019 June.

Table 2. Descriptive statistics of MBS items claimed (pooled data)

	Mean	SD	Median	Min	Max
Panel A: Doctor characteristics (n= 2,428,932)					
GPs	0.536	0.499	1	0	1
Specialists	0.464	0.499	0	0	1
Male	0.655	0.475	1	0	1
Female	0.345	0.475	0	0	1
Born at/after 1960	0.461	0.498	0	0	1
Born before1960	0.539	0.498	1	0	1
Outside major cities	0.411	0.492	0	0	1
Major city	0.589	0.492	1	0	1
Qualification from Australia	0.786	0.410	1	0	1
Qualification from overseas	0.214	0.410	0	0	1
Panel B: Items claimed per month per doctor (all items) (n= 2,428,932)					
Share of the item for each doctor (2015)	0.072	0.162	0.012	0	1
Total medical expenditure ($)	1,479	3,946	325.6	0	212,184
Total medical expenditure by Medicare ($)	1,064	2,957	246.7	0	183,387
Number of claims per item	19.37	140.4	3	1	19,022
Proportion of bulk-billed claims	0.619	0.449	1	0	1
Average fee ($)	185	349.1	94.07	0	10,612
Panel C1: Items claimed per month per doctor (items affected any time n=167,543)					
Share of the item for each doctor (2015)	0.155	0.249	.022	0	1
Total medical expenditure ($)	2,860	5,739	659.8	0	145,719
Total medical expenditure by Medicare ($)	1,626	2,903	451.7	0	44,105
Number of claims per item	24.50	51.71	4	1	835
Proportion of bulk-billed claims	0.464	0.444	0.333	0	1
Average fee ($)	252.9	504.6	83.29	0	9,999
Types of changes recommended by Taskforce (n=26,834, affected claims *after* implementation)					
Amend	0.864	0.343	1	0	1
Delete	0.039	0.194	0	0	1
Add	0.097	0.296	0	0	1
Panel C2: Items claimed per month per doctor (items not affected at all n=2,261,389)					
Share of the item for each doctor (2015)	0.065	0.151	0.012	0	1
Total medical expenditure ($)	1,377	3,760	312.7	0	212,184
Total medical expenditure by Medicare ($)	1,022	2,957	238.7	0	183,387
Number of claims per item	18.99	144.8	2	1	19,022
Proportion of bulk-billed claims	0.630	0.447	1	0	1
Average fee ($)	180.2	334.2	94.83	0	10,612

Notes: The unit of observation in the analysis is item-doctor-month. Variables for doctor's characteristics are from MABEL and are based on the first response for each doctor as MABEL information is only available before 2019.

Table 3. Before and after comparison of items that were amended, deleted or added (n= number of claims)

	12 months before	12 months after
	Amend (n=17,870)	Amend (n=15,165)
Total medical expenditure ($)	3,225	3,601
Total medical expenditure by Medicare ($)	1,787	1,990
Number of claims per item	27.80	29.59
Proportion of bulk-billed claims	0.458	0.426
Average fee ($)	267.3	290.9
	Deletion (n=826)	Deletion
Total medical expenditure ($)	6,050	n/a
Total medical expenditure by Medicare ($)	3,366	n/a
Number of claims per item	12.05	n/a
Proportion of bulk-billed claims	0.083	n/a
Average fee ($)	508	n/a
	Addition	Addition (n=1,867)
Total medical expenditure ($)	n/a	1,687
Total medical expenditure by Medicare ($)	n/a	906.4
Number of claims per item	n/a	7.446
Proportion of bulk-billed claims	n/a	0.571
Average fee ($)	n/a	379.3

Notes: Group id is item-doctor. In the "12 months after" column, we include the 12 months *one month* after the implementation occurred. For instance, if the change happened in July 2018, then we include the observations from August 2018. For "amended" and "added" changes that occurred in July and November 2018, we do not have the full year information as our data ended in June 2019. Therefore, we estimate the annual information for those particular items by multiplying them up to 12 months (e.g., expenditure*12/(months available).

Table 4. Number and percentage of items affected by doctor or item characteristics

	Never-affected	Affected any time of the analysis	Total
Item group			
Consultation items	1,094,337 (94.6)	62,043 (5.37)	1,156,380
Diagnostic Tests and Pathology	362,486 (89.1)	41,481 (10.9)	407,967
Obstetrics, Anesthetics, Operations and others	806,566 (93.0)	61,019 (7.05)	865,585
Doctor characteristics			
GPs	1,265,376 (97.2)	36,850 (2.83)	1,302,226
Non-GP specialists	996,013 (88.4)	130,693 (11.6)	1,126,706
Male	1,477,134 (92.8)	113,899 (7.16)	1,591,033
Female	784,255 (93.6)	53,644 (6.40)	837,899
Born at/after 1960	1,040,176 (93.0)	78,392 (7.01)	1,118,568
Born before1960	1,222,213 (93.2)	89,151 (6.80)	1,310,364
Outside major cities	934,419 (93.7)	63,178 (6.33)	997,597
Major city	1,327,970 (92.7)	104,365 (7.29)	1,431,335
Trained in Australia	1,778,738 (93.2)	130,068 (6.81)	1,908,806
Trained overseas	482,651 (92.8)	37,475 (7.20)	520,126

Notes: Numbers in parentheses represent the percentage, which sums into 100 % for each row.

Table 5. Main results: Average Treatment Effect on the Treated (n=2,428,932)

	Ln (Total medical exp.) [1]	Ln (Total exp. by Medicare) [2]	Ln (No. of claims per item) [3]	Proportion bulk billed [4]	Ln (Average fee) [5]
Overall effect					
$ATT\ (\theta_{sel}^{O})$	-0.038	-0.045	-0.034	0.002	-0.004
s.e.	0.040	0.036	0.035	0.010	0.009
Group-time effects					
ATT (2017/Nov)	-0.043	-0.044	-0.035	0.008	-0.009
s.e.	0.040	0.039	0.039	0.009	0.008
ATT (2018/Mar)	-0.064	-0.097	-0.047	-0.034	-0.017
s.e.	0.170	0.181	0.181	0.020	0.019
ATT (2018/Nov)	0.046	-0.023	-0.020	-0.043	0.066
s.e.	0.171	0.127	0.120	0.036	0.060

Notes: The control group is items that were never treated. The doubly robust method uses time-invariant covariates (i.e., year of birth before 1960, working in a major city, receiving qualification from overseas, indicators of being a GP, male or consultation items, the share of each item based on 2015-claims per doctor, MBS item group). We use bootstrap for standard errors that are in parentheses. The time variable is the year-month level. Group id is doctor-item. * = p value<0.10; ** = p value<0.05; *** = p value<0.01

Table 6. Heterogeneity of ATT based on doctor's characteristics

	Ln (Total medical exp.)	Ln (Total exp. by Medicare)	Ln (No. of claims per item)	Proportion of bulk-billed	Ln (Average fee)
	[1]	[2]	[3]	[4]	[5]
Panel A					
GP (n=1,302,226)					
$ATT\ (\theta^{O}_{sel})$	-0.170	-0.152	-0.271	-0.014	0.101***
s.e.	0.222	0.216	0.227	0.067	0.038
Specialists (n=1,126,706)					
$ATT\ (\theta^{O}_{sel})$	-0.030	-0.039	-0.028	0.001	-0.002
s.e.	0.039	0.035	0.035	0.009	0.008
Panel B					
Born before 1960 (n=1,310,364)					
$ATT\ (\theta^{O}_{sel})$	-0.024	-0.037	-0.017	-0.013	-0.007
s.e.	0.063	0.058	0.059	0.011	0.015
Born at/after 1960 (n=1,118,568)					
$ATT\ (\theta^{O}_{sel})$	-0.044	-0.044	-0.044	0.016	-3e-04
s.e.	0.056	0.053	0.052	0.018	0.014
Panel C					
Work in major cities (n=1,431,335)					
$ATT\ (\theta^{O}_{sel})$	-0.077	-0.073	-0.069	0.011	-0.007
s.e.	0.050	0.050	0.049	0.011	0.009
Work outside major cities (n=997,597)					
$ATT\ (\theta^{O}_{sel})$	0.066	0.037	0.037	-0.024	0.012
s.e.	0.074	0.065	0.064	0.018	0.018
Panel D					
Qualified overseas (n=520,126)					
$ATT\ (\theta^{O}_{sel})$	-0.091	-0.108	-0.069	0.006	-0.022
s.e.	0.101	0.109	0.113	0.018	0.024
Qualifed in Australia (n=1,908,806)					
$ATT\ (\theta^{O}_{sel})$	-0.032	-0.035	-0.031	-0.001	-9e-04
s.e.	0.045	0.042	0.043	0.012	0.012
Panel E					
Female (n=837,899)					
$ATT\ (\theta^{O}_{sel})$	0.006	-0.095	-0.071	0.013	-0.025
s.e.	0.008	0.070	0.069	0.023	0.026
Male (n=1,591,033)					
$ATT\ (\theta^{O}_{sel})$	9e-04	-0.007	-0.005	-0.004	0.006
s.e.	0.052	0.051	0.046	0.009	0.008

Notes: The control group we use here is never-treated. The doubly robust method uses time-invariant covariates (i.e., year of birth before 1960, working in a major city, receiving qualification from overseas, indicators of being a GP, male or consultation items, the share of each item based on 2015-claims per doctor, and MBS item group). We use bootstrap for standard errors that are in parentheses. The time variable is the year-month level. Group id is doctor-item. * = p value<0.10; ** = p value<0.05; *** = p value<0.01

Table 7. Robustness Checks

	Ln (Total medical exp.) [1]	Ln (Total exp. by Medicare) [2]	Ln (No. of claims per item) [3]	Proportion of bulkbilled [4]	Ln (Average fee) [5]
1. Control group changed: using the not-yet treated group as a control group (n=2,428,932)					
$ATT\ (\theta^{O}_{sel})$	-0.037	-0.045	-0.045	0.002	-0.003
s.e.	0.040	0.038	0.041	0.010	0.009
2. Change estimand from Doubly Robust to IPW only (n=2,428,932)					
$ATT\ (\theta^{O}_{sel})$	-0.039	-0.045	-0.045	0.002	-0.004
s.e.	0.041	0.040	0.038	0.009	0.009
3. Change estimand from Doubly Robust to Outcome Regression only (n=2,428,932)					
$ATT\ (\theta^{O}_{sel})$	-0.004	-0.039	-0.048	-0.048	7e-4
s.e.	0.009	0.037	0.035	0.032	0.009
4. **Without any covariates (n=2,428,932)**					
$ATT\ (\theta^{O}_{sel})$	-0.084**	-0.113**	-0.113**	-0.004	0.007
s.e.	0.030	0.028	0.029	0.008	0.006
5. Exclusion of four doctors who made more than 3,000 claims per item-month-doc (n=2,394,285)					
$ATT\ (\theta^{O}_{sel})$	-0.038	-0.045	-0.034	0.002	-0.004
s.e.	0.038	0.039	0.039	0.010	0.009
6. **Exclusion of Pathologists and Radiologists (n=2,266,423)**					
$ATT\ (\theta^{O}_{sel})$	-0.035	-0.043	-0.031	8e-04	-0.004
s.e.	0.043	0.037	0.037	0.010	0.010
7. **Including years from August 2014 (n=1,542,102)**					
$ATT\ (\theta^{O}_{sel})$	-0.050	-0.063	-0.043	0.005	-0.007
s.e.	0.033	0.034	0.032	0.007	0.007

Notes: We control time-invariant covariates (i.e., year of birth before 1960, working in a major city, receiving qualification from overseas, indicators of being a GP or male, the share of each item based on 2015 all claims per doctor, and MBS item group). The time variable is the year-month level. Group id is doctor-item. * = p value<0.10; ** = p value<0.05; *** = p value<0.01

Appendix

Creation of dataset of MBS item changes arising from that MBS Review Taskforce that was implemented by the Department of Health.

We manually created a dataset of all changes recommended by the MBS Review Taskforce and implemented by the Department of Health. This includes detailed information of the changed item, including the group that this item is in, the date of the implemented change, and the type of changes. This dataset is based on several sources.

There was no detailed published comprehensive list of all changes recommended and implemented. We contacted the Department of Health, and they advised we use individual factsheets published on the MBS Review website. Each factsheet mentions if the MBS Review Taskforce recommended the change for each item.[10] We included all changes implemented by the Department of Health in the factsheet in our dataset.[11] Second, we also cross-checked this with information from the MBS website that periodically reports *all* changes to the MBS from January 2016 to June 2019.[12] From there, we selected the subset of items that were implemented *only* based on the MBS Review Taskforce recommendations. To do this, we first identify the items from the MBS webpages (e.g., a summary of changes for November 2018) and whether there is any note that indicates whether the changes were a result of the MBS Review.[13]

Finally, since the explanatory notes under each MBS news may not identify all changes recommended by the Taskforce, we also cross-checked with the broad implementation timeline of the MBS Review Taskforce recommendations from 2016 to 2019 on the Australian Government - Department of Health website. This timeline presents a year of implementation

10 For 2017 changes: http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Factsheet-Previous. For 2018 changes: http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Factsheet-2018.

11 For example, factsheet on "Ear, Nose, and Throat" shows that in November 2017, there were four changes made to these items: 41674, 41789, 41793, and 41801. These changes are the amendments in tonsillectomy and adenoidectomy items to avoid inappropriate billing. Therefore, in our data, these four changes are included, and noted as "amended" items.

12 http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/news

13 In November 2017, for example, the note on the website says "the MBS Quantitative Computed Tomography (QCT) items (12309 and 12318) are removed from the MBS following review under MBS Review Taskforce processes, on the basis that QCT provides lower value care in comparison to Dual Energy X-ray Absorptiometry." Based on this, we include these two items in the dataset as "deleted" items. See below:
http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/news-2017-11-01-latest-news-November

for each item change recommended by the Taskforce.[14] For example, in 2017, changes in five categories by the Taskforce were implemented: (1) obstetrics, (2) ear, nose, and throat, (3) diagnostic imaging, (4) principles and rules, and (5) gastroenterology. These did not mention specific item numbers but broad item groups. We then map these to changes from publications based on MBS Review Taskforce recommendations and to see if we missed any implemented changes in our dataset from the first two steps.[15]

14 The implementation timing of each recommendation may vary due to complexity and impacts on stakeholders. The specific month/date for implementation of recommendations is not available in this timeline. The timeline is available here: https://www.health.gov.au/initiatives-and-programs/mbs-review

15 https://www.health.gov.au/resources/collections/mbs-review-final-taskforce-reports-findings-and-recommendations

Table A1. Comparison between characteristics of MABEL GPs & specialists consenting to MBS linkage with the population of GPs & specialists in 2015

	MABEL GPs & specialists who consented to linkage to MBS. Total = 2,216 (n, %)	Population Total = 52,299 (n, %)
GP	1,113 (50.2)	26,635 (50.8)
Specialist	1,103 (49.8)	25,759 (49.2)
Male	1,303 (58.8)	34,262 (65.4)
Female	913 (41.2)	18,132 (34.6)
Born in/after 1960 (n = 2,212)	1,254 (56.7)	31,806 (61.7)
Born before 1960 (n = 2,212)	958 (43.3)	19,715 (38.3))
Outside major city (ASGS)	647 (29.2)	10,687 (20.8)
Major City (ASGS)	1,569 (70.8)	40,766 (79.2)
Qualification from Australia (n = 2,164)	1,676 (77.4)	33,179 (66.4)
Qualification from overseas (n = 2,164)	488 (22.6)	16,826 (33.6)

Source: MABEL-MBS linked data, and Medical Directory of Australia, doctors in clinical practice.

Table A2. Number and distribution of claims in 2015 (GPs)

	Sample (n, %)	National (n, %)
A1 General Practitioner Attendances	3,121,824 (86.6)	113,551,153 (82.2)
A15 Multidisciplinary Care Plans and Case Conferences	204,051 (5.7)	7,107,882 (5.2)
A20 GP Mental Health Treatment	84,689 (2.6)	3,007,274 (2.2)
A22 GP after-hours attendances to which no other item applies	68,950 (1.9)	8,770,840 (6.4)
M12 Services provided by a Practice Nurse	55,427 (1.5)	1,557,115 (1.1)
A14 Health Assessments	30,080 (0.8)	998,881 (0.7)
A18 GP attendance associated with PIP incentive payments	17,633 (0.5)	404,238 (0.3)
A7 Acupuncture and non-specialists practitioner items	10,412 (0.3)	518,687 (0.4)
A11 After Hours	6,017 (0.2)	1,691,392 (1.2)
A17 Domiciliary Medication Management Review	5,339 (0.2)	127,945 (0.1)
A23 Other non-referred after-hours attendances to which no other item applies	712 (<0.1)	318,577 (0.2)
A6 Group Therapy (other than by psychiatrist)	669 (<0.1)	12,040 (<0.1)
A19 Other non-referred attendance associated with PIP incentive payments	49 (<0.1)	7,375 (<0.1)
Total	3,605,852	138,073,399

Source: MABEL-MBS linked data, and Medical Directory of Australia, doctors in clinical practice. Excludes diagnostic imaging and pathology as the national data do not distinguish which type of doctors made the claim (see Table A4).

Table A3. Number and distribution of claims in 2015 (Non-GP specialists)

	Sample (n, %)	National (n, %)
A4 Consultant Physician (other than psychiatrist)	374,120 (24.8)	12,203,499 (22.0)
A3 Specialist Attendances	348,541 (23.1)	11,605,214 (20.9)
T8 Surgical Operations	239,481 (15.9)	9,977,475 (18.0)
T1 Miscellaneous Therapeutic Procedures	103,126 (6.8)	3,075,373 (5.5)
T10 Relative Value Guide for Anaesthesia	92,290 (6.1)	2,879,802 (5.2)
T6 Anaesthetics	84,593 (5.6)	2,568,242 (4.6)
T4 Obstetrics	80,918 (5.4)	1,942,613 (3.5)
A8 Consultant Psychiatrist	58,774 (3.9)	2,290,696 (4.1)
T2 Radiation Oncology	57,478 (3.8)	1,988,754 (3.6)
A2 Other non-referred	32,528 (2.2)	5,567,290 (10.0)
A24 Pain and Palliative Medicine	8,522 (0.6)	159,997 (0.3)
T9 Assistance at Operations	7,262 (0.5)	490,081 (0.9)
T7 Regional or Field Nerve Blocks	5,998 (0.4)	338,323 (0.6)
T11 Botulinum Toxin Injections	4,472 (0.3)	54,570 (0.1)
A26 Neurosurgery Attendances to which no other item applies	3,958 (0.3)	225,898 (0.4)
A21 Medical Practitioner (Emergency Dept, private hospital)	3,128 (0.2)	96,162 (0.2)
A28 Geriatric Medicine	1,771 (0.1)	47,286 (0.1)
A30 Medical Practitioner video conferencing	1,145 (0.1)	34,901 (0.1)
A29 Early Intervention Services for Children	606 (<0.1)	10,257 (<0.1)
A5 Prolonged	406 (<0.1)	18,008 (<0.1)
T3 Therapeutic Nuclear Medicine	121 (<0.1)	4,014 (<0.1)
A27 Pregnancy Support Counselling	88 (<0.1)	8,682 (<0.1)
A13 Public Health Physician Attendances	44 (<0.1)	6,965 (<0.1)
A9 Contact Lenses	34 (<0.1)	369 (<0.1)
C2 Oral Surgical Services	8 (<0.1)	547 (<0.1)
	1,509,412	55,595,018

Source: MABEL-MBS linked data, and Medical Directory of Australia, doctors in clinical practice. Excludes diagnostic imaging and pathology as the national data do not distinguish which type of doctors made the claim (see Table A4).

Table A4. Number and distribution of claims in 2015 (Diagnostic imaging and pathology)

	Sample (n, %)	**National (n, %)**
P2 Chemical	302,773 (24.0)	47,544,209 (28.2)
P10 Patient Episode Initiation	271,396 (21.4)	39,498,358 (23.4)
D1 Miscellaneous Diagnostic Procedures and Investigations	127,370 (10.0)	6,854,308 (4.1)
P1 Haematology	126,982 (10.0)	17,387,509 (10.3)
P3 Microbiology	113,059 (8.9)	14,832,803 (8.8)
I3 Diagnostic Radiology	73,389 (5.8)	10,460,056 (6.2)
I1 Ultrasound	72,715 (5.7)	9,238,320 (5.5)
P5 Tissue Pathology	45,380 (3.6)	3,342,966 (2.0)
P4 Immunology	37,161 (2.9)	4,001,601 (2.4)
I2 Computerised Tomography	22,699 (1.8)	2,834,594 (1.7)
P11 Specimen Referred	15,469 (1.2)	485,250 (0.3)
I4 Nuclear Medicine Imaging	15,467 (1.2)	660,528 (0.4)
P6 Cytopathology	14,356 (1.1)	2,070,342 (1.2)
P9 Simple Basic Tests	13,252 (1.0)	652,976 (0.4)
I5 Magnetic resonance imaging	11,301 (0.9)	998,301 (0.6)
P8 Infertility and Pregnancy Tests	4,331 (0.3)	711,224 (0.4)
P7 Cytogenetics	1,926 (0.2)	264,706 (0.2)
D2 Nuclear Medicine (non-imaging)	39 (<0.1)	6,880,165 (4.1)
Total	1,269,065	168,718,216

Source: MABEL-MBS linked data, and Medical Directory of Australia, doctors in clinical practice.

THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE
APPLIED ECONOMIC & SOCIAL RESEARCH

60 YEARS IMPACT