



**HAL**  
open science

# How does the quality of care for type 2 diabetic patients benefit from GPs-nurses' teamwork? A staggered difference-in-differences design based on a French pilot program

Julie Gilles de La Londe, Anissa Afrité, Julien Mousques

## ► To cite this version:

Julie Gilles de La Londe, Anissa Afrité, Julien Mousques. How does the quality of care for type 2 diabetic patients benefit from GPs-nurses' teamwork? A staggered difference-in-differences design based on a French pilot program. *International Journal of Health Economics and Management*, 2023, 10.1007/s10754-023-09354-z . hal-04116955

**HAL Id: hal-04116955**

**<https://hal.science/hal-04116955>**

Submitted on 21 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

**How does the quality of care for type 2 diabetic patients benefit from GPs-nurses' teamwork? A staggered difference-in-differences design based on a French pilot program.**

Julie Gilles de La Londe<sup>b</sup>, Anissa Afrite<sup>c</sup>, Julien Mousquès<sup>a,c</sup>

<sup>a</sup> *Ecole des Hautes Etudes en Santé Publique (EHESP), ARENES (UMR 6051), 15 Av. du Professeur Léon Bernard, 35043 Rennes, France.*

<sup>b</sup> *Université de Paris, Département de Médecine Générale, 16 rue Henri Huchard, 75018 Paris, France.*

<sup>c</sup> *Institute for Research and Information in Health Economics (IRDES), 21 Rue des Ardennes, 75019 Paris, France.*

- Corresponding author: [mousques@irdes.fr](mailto:mousques@irdes.fr)
- ORCID number: <https://orcid.org/0000-0003-0844-6309>

**PRE PROOF VERSION**

**Published in :**

**Gilles de la Londe, J., Afrite, A. & Mousquès, J. How does the quality of care for type 2 diabetic patients benefit from GPs-nurses' teamwork? A staggered difference-in-differences design based on a French pilot program. Int J Health Econ Manag. (2023). <https://doi.org/10.1007/s10754-023-09354-z>**

## **ABSTRACT**

In many countries, policies have explicitly encouraged primary care teams and inter-professional cooperation and skill mix, as a way to improve both productive efficiency gains and quality improvement. France faces barriers to developing team working as well as new and more advanced roles for health care professionals including nurses.

We aim to estimate the impact of a national pilot experiment of teamwork between General Practitioners (GPs) and Advance Practitioners Nurses (APN) – who substitute and complement GPs – on yearly quality of care process indicators for type two diabetes patients (T2DP). Implemented by a not-for-profit meso-tier organisation and supported by the Ministry of Health, the pilot relied on the voluntary enrolment of newly GPs from 2012-2015; the staffing and training of APNs; skill mixing and new remuneration schemes.

We use latent-response formulation models, control for endogeneity and selection bias by using controlled before-after and quasi-experimental design combining coarsened exact matching – prior to the treatment, at both GPs (435 treated vs 973 control) and T2DP levels –, with intention to treat (ITT; 18,310 in each group) and per protocol (PP, 2943 in each group) perspectives, as well as difference-in-differences (DID) estimates on balanced panel claims data from the National Health Insurance Fund linked to clinical data over the period 2010-2017.

We show evidence of a positive and significant positive impact for T2DP followed-up by newly enrolled GPs in the pilot compared to the pretreatment period and the control group. The effect magnitudes were larger for PP than for ITT subsamples.

Keywords and Mesh terms: General Practitioners; Nurse's Role; Teamwork; Diabetes Mellitus; Controlled Before-After Studies; Difference-in-differences.

JEL Classification: I18 I11 J24 J44 C23 C21

## **Acknowledgments**

We would like to acknowledge the Directorate of Strategy, Studies and Statistics (DSES) of the National Health Insurance (NHI) and the not-for-profit organization *Asalée* for having facilitated accessibility to the dataset and especially the following people: Claude Gissot, Frédéric Bousquet, Pierre Bergman, Mehdi Gabbas, and Brice Dufresne (DSES) and Jean Gautier and Amaury Derville (Asalée). We are also grateful to Denis Raynaud for his support and Fabien Daniel and Charlie Menard for early involvement

on data collection and management. This research was undertaken within the framework program of the Institute for Research and Information in Health Economics (IRDES). Finally, we warmly thank all the staff and health care professionals involved for their support.

**Ethical approval:** Decree 2014-1181 of 13 October 2014 from the French Council of State.

## Introduction

While health policy of most OECD countries faces the challenge of strengthening and adapting their primary health care (PHC) delivery system, through organizational, technological and incentives change, France seems to struggle with the issue of high quality of care for chronic patient including type two diabetes patients and ranks particularly low regarding the proportion of people suffering from diabetes correctly followed-up (OECD, 2020).

The prevalence of treated diabetic patients in the French population covered by the general healthcare scheme in 2019 was 5.2 % on average (3.5 million) with huge geographical disparities and a moderate but constant yearly increase in the prevalence rate – approximately +0.9% each year for males and +0.4 % each year for females since 2010. Diabetes has the highest prevalence among all 100 %-covered chronic conditions. In the French health system, 87 % of patients with type 2 diabetes mellitus (T2DM) are treated in primary care by a general practitioner (GP). In the past fifteen years, French State and National Health Insurance (NHI) focused on four main policy levers to improve the management of T2DM patients (Chevreul et al., 2020): 1) guideline diffusion by the High Authority in Health (HAS); 2) pay-for-performance (P4P) payment mechanisms for GPs to increase GPs' adherence to guidelines; 3) support of diabetes networks between primary care and secondary care providers; and 4) implementation of a disease management program directly led by the NHI (so-called *Sophia*) for management, patient education and counseling. However, the impact of these numerous and large policies on the quality of care remains weaker than expected. The gap with the target of 80 % of T2DM patients correctly followed-up according to a set of process indicators, fixed in the 2004 Public Health Law, is still rather large for numerous indicators with only two notable exceptions, i.e., annual creatinemia and lipid check-up, for which 89.2 % and 76.9 % of T2DM patients, respectively, were tested yearly in 2019<sup>1</sup>. Nevertheless, all the other 2019 figures raised questions regarding the quality of care and services delivered for T2DM patients and the lack of dramatic improvement during the last 10 years (Fosse-Edorh & Mandereau-Bruno, 2015). Only 56 % of

---

<sup>1</sup> See for further details: <https://geodes.santepubliquefrance.fr/#c=home> and <https://www.santepubliquefrance.fr/les-actualites/2020/journee-mondiale-du-diabete-14-novembre-2020>

the T2DM patients had 3 HbA1c tests every year; 42 % an annual microalbuminuria test, 64 % an eye examination, 37.5 % a cardiovascular visit, and 38.3 % a dental visit. Macrovascular complications represent the leading cause of morbidity. They remain higher in France than in other OECD countries and their burden is expected to grow in the coming years (OECD, 2015). To date, the early prevention and management of T2DM patients within PHC remain a challenge. France is seeking regulations, organizational and technological change, and right incentives to develop comprehensive PHC notably in order to deliver high quality of care for chronic patient at the general practice level. Fostering interprofessional cooperation, task shifting and advanced roles for health care professionals including nurses might be part of the solution (Freund et al., 2015; de Bont et al., 2016) but still is a challenge for France (Kringos et al., 2013; Maier & Aiken, 2016; Maier et al., 2017). Most French primary healthcare professionals including GPs and nurses are self-employed, practicing in private and monodisciplinary practices, mostly paid via fees-for-services (FFS). Multi or interprofessional Primary Care Teams (PCT) are less developed in France compared to other countries, 61 % of GPs practice in group but less than half of them with paramedics including nurses (Chaput et al., 2019). Professional boundaries remain strong, and GPs are still reluctant to delegate to nurses (Massin et al., 2014), but the current circumstances provide great opportunities to enable new and advanced nursing roles. First, the proportion of GPs and nurses working in PCTs has drastically increased with more than 1,900 PCTs in 2021 compared to fewer than 20 in 2008 (Chevillard et al., 2019). Second, GPs have become a scarcer resource than nurses. This raises questions regarding the increasing percentage of the population living in medically underserved areas with difficult access to primary care, from 4% in 2005 to 18 % in 2017 (Legendre, 2020). Third, the Bologna process in Europe enables nurses to obtain master's degrees in nursing (Collins & Hewer, 2014). Finally, an advanced practice nurse (APN) statutory order was passed with the healthcare law in 2018 (Debout, 2018).

In this context, the French government financially supported, during the period 2010–2018, the extension of a pilot program encouraging teamwork between GPs and nurses called “*Action de santé libérale en équipe*” (*Asalée*), that is, teamwork for self-employed health care professionals (Fournier et al., 2018). Nurses with advance or extended role – called Advance Practitioners Nurses (APN) for the remainder of the paper – enrolled in this pilot program were trained, hired

and paid by the eponymous nonprofit organization. Pilot program funding covered investment and operating costs, the wages of APNs and pay-for-coordination for the GPs. The pilot program allowed the GPs to cooperate with APNs, who carried out activities that are usually, but not systematically, undertaken by GPs (screening, health education, technical procedures). Based on the national extension of this pilot, we aim to identify the effect of teamwork between GPs and APNs on yearly quality of care process indicators for type two diabetes patients (T2DP), all the more so if we know that there is no other primary care organization than the *Asalée* pilot involving APN in such a way. The collaboration between GPs and nurses seems all the more effective as the intensity and the quality of the interaction of the GP-nurse pairs is strong. It creates bonds between cure and care through a holistic approach (Barnett, 2018). The literature agrees on the rather positive impact of cooperation and task shifting between GPs and nurses (Martínez-González et al., 2015; Tsiachristas et al., 2015; Laurant et al., 2018) but uncertainty remains regarding the impact of doctor-nurse's skill mix on other care processes due to low evidence. We use the quasi-experimental framework provided by the pilot, with a large national extension that led many GPs and APNs to enroll the pilot between 2012-2015. Using balanced panel data resulting from the pairing of National Health Insurance (NHI) claims data with clinical data from the pilot, every year over the period 2000-2017, we provide empirical evidence of the positive impact of the *Asalee* APNs' pilot on the quality of care provided for type 2 diabetes patients. To control for endogeneity and selection bias at both the GP and patient levels, we used both the properties of coarsened exact matchings prior to the treatment, to identify similar controlled groups based on observables, and difference-in-differences (DiD) estimates. It should be noted that our quasi-experimental design does not solve all questions regarding the generalizability of our results since the effect of unobservable variables on the treated and control groups could potentially change over time. Also, to take into account of our specific DiD setup, with multiple time periods and variation in treatment exposure or timing, as well as of potential heterogeneous treatment effects, we used a staggered adoption DiD proposed by Callaway and Sant'Anna based on event study DiD specifications (Callaway and Sant'Anna, 2021). Finally, we focus on per protocol patient (PP, 2943 in each group), regarding the effect of the treatment on the treated, but considering intention to treat perspective (ITT; 18,310 in each group) as robustness check analysis in order to test if the impact could be attributed to GPs APNs

cooperation and/or to GPs general improvement of practice patterns regarding T2DP management.

The remain of the paper is organized as follows. We present the institutional setting as well as the main empirical and theoretical effects that could be expected from cooperation between GPs and APNs in section 2. Section 3 describes the data collection, management, the design and empirical strategy. Section 4 highlights the results and the robustness checks and Section 5 discusses them.

## **Background**

### **Institutional setting**

The French PHC is essentially publicly funded, the National Health Insurance (NHI) finances two-thirds of the PHC expenditures and the remaining is paid directly by households and/or complementary health insurance companies. PHC providers are mainly self-employed professionals and practitioners (including specialists and paramedics) are mostly paid by fee-for-service (FFS) up to 78% for GPs in 2019 (DREES, 2020). It is estimated that 61% of GPs practice in group in 2019 (Chaput et al., 2019) and if multi-professional primary care teams (PCTs) remained a marginal way of practice, it increases dramatically with more than 1,900 PCTs in 2020 compared to fewer than 20 in 2008 (Chevillard et al., 2019).

Initially, the *Asalée* pilot began in 2004 in 3 practices (12 GPs and 3 APNs) within a county council and was supported both by a physician self-employed regional organization and by public funds. It was extended to 18 practices (41 GPs and 8 APNs) from 2007-2009 and then extended again to include 293 additional practices (780 GPs and 220 APNs) between 2010 and 2015 at the national level (Fournier et al., 2018). Currently, the *Asalée* pilot consists of a not-for-profit meso-tier organization, supported by the state (Social Security Department) and the NHI, who enrolled GPs and hired APNs to collaborate with them. The state and the NHI defined ex-ante the FTE of newly enrolled ANs and, indirectly, the number of GPs enrolled for a given period, and explicitly asked for mainly rural or disadvantaged urban areas to be targeted regarding GP density. Public financing partly or wholly covered the investment and operational

costs, especially nurses' salaries, and payments for coordination made to GPs (in addition to FFS and other prospective fees). Note that GPs' participation in the *Asalée* pilot was and has remained voluntary, but it implies the participation of the whole GPs' practice, APNs are selected, hired, trained, and paid by the nonprofit organization *Asalée* to work in one or more practices with one or several GPs. One could expect that this process results in large selection bias. Fournier and colleagues (2018), based on a comprehensive sociological analysis focusing on GP and APN trajectories, expectations, working conditions, and cooperation, show that ANs have high and relatively homogenous motivations concerning the development of therapeutic patient education, teamwork with GPs, professional autonomy, and advanced practices. While during the pilot's first years, GPs who enrolled were relatively homogeneous, thanks to the persuasive efforts of GPs who were also part of the medical union's network, recruitment has been less determined and controlled by the *Asalée* not-for-profit organization since the regional and national extensions. Therefore, for instance, if a GP who supports the project within a specific practice expresses high expectations, the other GPs could be either enthusiastic, neutral, or even reluctant or quite skeptical. This led to the recruitment of heterogeneous GPs, some very aware and well informed, and others who had no real understanding of the pilot program. Our design and empirical strategy combine exact matching and difference-in-differences estimations using panel data to take selection issues at the GP level into account as much as we can.

The pilot enabled APNs to obtain advanced roles in the management of T2DM patients. These roles included (1) identifying patients at risk for T2DM, updating T2DM patients' medical records (follow-up laboratory and clinical tests), and setting electronic reminders in the electronic medical records; (2) performing EKG and foot exams based on a physician order; and (3) patient education and counseling on GP's referral.

Focusing on the *Asalée* pilot program, patient education enables T2DM patients to improve their knowledge, skills and confidence; to take increasing control of their health condition and to integrate effective self-management on a daily basis (Fournier et al., 2018). Qualitative works on *Asalée* also showed that the nature and intensity of the collaboration and the balance between delegated tasks from the GPs to the nurses (performing EKG and foot exam), which ends with substitution, and additional tasks for nurses to deliver new care and services to the patients

(patient education), which ends with complementarity or supplementation, are heterogeneous between APN-GP pairs (Afrite et al., 2019).

### **Theoretical and empirical related literature**

From a theoretical point of view, the *Asalée* pilot program could be qualified as integration between GPs and APNs (i.e., teamwork, skill mixing, and cooperation) that can generate – in addition to efficiency gains – quality improvement under certain circumstances and depending on the nature of the cooperation (Kernick & Scott, 2002; Nicolson & Propper, 2011). Strictly speaking and following Maier and colleagues (Maier et al., 2016a) skill mix includes change in “skills, competencies, roles or tasks within and across health professionals and health workers (including community-based workers, peers, informal caregivers) and/or teams”. It could end both with re-allocation or delegation (substitution) of tasks and/or addition of new tasks/roles (supplementation), with or without change in teamworking. Then one could expect productive efficiency gains from substitution and quality improvements from supplementation.

Three systematic reviews reviewed the evidence on the effectiveness of physician–nurse task shifting in primary care out of 53 studies conducted between 1994 and 2012, and none of the studies was French (Martínez-González et al., 2015; Tsiachristas et al., 2015; Laurant et al., 2018). The literature agrees on the impact of cooperation and task shifting between GPs and nurses when nurses substitute or supplement the physician workforce. A recent meta-analysis by Laurant et al. (Laurant et al., 2018) and a review by Maier et al. (Maier et al., 2022b) showed that for some chronic conditions, trained nurses, compared to primary care physicians, likely provide equal outcomes in terms of mortality but possibly better outcomes in terms of blood pressure and patient satisfaction. That being said uncertainty remains regarding the impact of doctor-nurse’s skill mix on other care processes due to low evidence. Four studies specifically evaluated the impact of task shifting from GPs to specialized nurses (SNs) or advanced practice nurses (APNs) on the quality of care and clinical outcomes of T2DM patients. They all had a smaller sample size than our study but showed small improvements in HbA1c and lipidic means (Taylor et al., 2005; Vrijhoef et al., 2001), a positive effect on the system of delivering medicines to patients with diabetes, a reduction in the number of insulin and oral hypoglycemic agent medical errors (Carey et al., 2008), and better satisfaction of T2DM patients when followed by

nurses rather than GPs (Houweling et al., 2011). It seemed in those studies that nurses helped patients handle their own situations. Furthermore, Lenz *et al.* (Lenz et al., 2002) showed that nurse practitioners were more likely to document BMI, urinalyses and HbA1c in medical records for T2DM patients. Further research could focus on this aspect and the cost effectiveness of task shifting for T2DM patients.

As part of a large evaluation policy program and following previous evaluation in the early stage of this pilot program (Mousquès et al., 2010), the aim of our study was to measure the impact of the *Asalée* national pilot program of teamwork between GPs and APNs on the quality of care delivered to T2DM patients regarding care processes.

## **Methods**

### **Data collection and study design**

From an initial sample of 612 treated GPs and 1,512 randomly selected control GPs, and based on a controlled before-after and quasi-experimental design and after several steps of data compilation and matching at both GP and T2DM patient levels, we compared the quality of the T2DM management of a sample of 435 treated GPs – enrolled in the pilot program between 2012 and 2015 – matched with 973 control GPs over the 2010-2017 period (see Fig 1). Depending on the intention to treat (ITT; *i.e.*, all T2DM patients are considered) or the per protocol (PP; *i.e.*, with restriction to T2DM patients enrolled within a specific diabetic protocol program run by APNs) perspective analysis, the corresponding T2DM patients subsample sizes were 18,310 and 2,943, respectively, for both treated and control groups.

The identification of GPs enrolled in the *Asalée* pilot and ITT and PP T2DM patients both relied on claims data from the National Health Insurance fund (NHI), namely, the National Health Data System (SNDS) and the National Inter-Schemes System of Health Care Professionals (SNIR-PS), linked to a registry and clinical dataset specific to the pilot and provided by the *Asalée* NGO. T2DM patients were identified using their unique registration to a specific GP and oral anti-diabetics (OAD) or insulin therapy treatment and/or hospitalization for diabetes and/or exemption from copayments due to their T2DM.

We observed all GPs and their T2DM patients' visits (e.g., GP and cardiologist), laboratory exams (e.g., Hba1c) or procedures (e.g., EKG) from 2010 to 2017.

To address selection bias at both the GP and patient levels and account for unobserved heterogeneity, we used both the properties of coarsened exact matching (CEM) to identify controlled groups and difference-in-differences (DiD) estimations using event study DiD estimations for staggered treatment design to estimate the impact of GP enrollment in the pilot program on T2DP follow-up.

### **Matching and control group**

We used a CEM to match first-treated GPs and second-treated T2DM patients with their controls in 2011 to reduce model dependence and exposure to misspecification regarding matching covariates (Iacus et al., 2011; Ho et al., 2007). CEM is particularly effective at finding “twins” and ensuring their comparability across the entire distribution of all covariates (King & Nielsen, 2019).

The following GP characteristics were used for the first CEM at the GP level:

- GP characteristics: age, sex, and geographical location (according to four living area taxonomy based on the following dimensions: urban-rural and overserved-underserved);
- Characteristics of the patients on the list: age structure (0-15 years, 16-59 years, 60-69 years, and 70 years and over), type of compulsory health insurance coverage (salaried workers, self-employed workers, and agricultural workers), percentage of patients with a long-term illness, and percentage of patients with compulsory health insurance coverage and free complementary health insurance coverage specific to low-income individuals.

The following T2DM patients’ characteristics were used for the CEM at T2DM patients’ level, with one treated for one control: age group (quartile), gender, GP location (quartile of the so-called localized potential accessibility (LPA) indicator (Barlet et al., 2012), a French application of a two-step floating catchment area method to measure the balance between the supply and demand in family medicine in each living area), type of compulsory health insurance coverage, free complementary health insurance coverage, category of drug therapy for T2DM that distinguishes oral therapy (first line, second line dual therapy, third line triple oral therapy, and others) and insulin.

**Fig. 1** Study sample flow chart

	Treated		Control	
	GPs	T2DP	GPs	T2DP
Data from National Health Data System ( <i>SNDS</i> )	612		62,172	
Data from National Health Data System ( <i>SNDS</i> ) and random controlled GPs restricted to maximum 3 controls to 1 treated	612		1,561	
Intention To Treat: restricted to GPs enrolled within the <i>Asalée</i> pilot from 2012-2015 and their T2DM patients, coarsened exact matching of treated and controlled GPs (restricted to maximum 3 controls to 1 treated) for 2010, and balanced panel over the 2010-2017 period	435	18,638	973	43,011
Intention To Treat: restricted to GPs enrolled within the <i>Asalée</i> pilot from 2012-2015 and their T2DM patients, coarsened exact matching of T2DM patients for 2010 ( <i>restricted to maximum 1 control to 1 treated</i> ), and balanced panel over the 2010-2017 period	435	18,310	973	18,310
Per Protocol: restricted to GPs enrolled within the <i>Asalée</i> pilot from 2011-2015, treated T2DM patients enrolled in a specific diabetic protocol program, coarsened exact matching of T2DM patients for 2010 ( <i>restricted to maximum 1 control to 1 treated</i> ), and balanced panel over the 2010-2017 period	435	2,943	739	2,943

## Empirical strategy

The difference-in-differences (DiD) approach at the patient level using balanced panel data of National Health Insurance claims from 2010-2017 was used to control for unobservable constants over time. We estimated the quality of T2DM management through process of care indicators, following the Donabedian classification applied to primary care (Campbell, 2003) based on robust evidence for T2DM management (OECD, 2015). These process indicators were different from *outcomes* in that they indicated whether an exam was conducted but not its value or result. The indicators that we considered comparing treated and control T2DPs were the following: glycosylated hemoglobin (at least three HbA1c per year: yes or no), microalbuminuria (one per year: yes or no), creatinemia (*idem*), lipid check-up (*idem*), funduscopy or a visit to an ophthalmologist (*idem*), and EKG or a visit to a cardiologist (*idem*). The treatment effect (*i.e.*, enrollment of self-employed GPs in the *Asalée* pilot) measured the differences between process indicators observed with and without intervention (Rubin, 1974; Bertrand et al., 2004; Imbens et al., 2009). We used DiD models to estimate the evolution of the process indicators, our outcome variables, for both treated and control T2DPs by assuming that unobserved (individual and temporal) heterogeneity was constant over time, which is the parallel trend assumption.

Considering the staggered implementation of the *Asalée* pilot, *i.e.* treatment timing varies across units, as well as the potential heterogeneous effects of the treatment across treatment cohorts and time, estimating DiD using two-way fixed effects (TWFE) specification could end with biased estimates (Goodman-Bacon, 2021; Sun et Abraham, 2021; Baker et al., 2022). Major concerns are the fact that when treatment is staggered, parallel assumption can't hold and some treated units will receive negative weight being considered as control. To take this into account, we rely here on the event-study DiD specification recently proposed by Callaway and Sant'Anna (2021) that avoid "bad comparisons" between earlier and later treated units, following Abadie's work (2005), and implemented with the CSDiD stata module (Rios-Avila et al., 2022). Other alternatives are available to deal with the "bad comparisons" biases of the TWFE models, also by changing the set of units that are compared in the treatment effect estimation (Baker et al., 2022). Our choice with the Callaway and Sant'Anna (CS) causal estimands is motivated by the following main reasons. First, they proposed unbiased and consistent group-time average

treatment effects on the treated (ATT) that allow us to analyze treatment effect heterogeneity between cohorts, at different times and with different lengths of treatment exposure. Second, the identification assumptions are easy to comply with knowing the way the *Asalée* pilot was rolled out and notably the irreversibility of the treatment, *i.e.* once a unit becomes treated it will remain treated in the next period, but also the two other key assumptions, limited treatment anticipation and conditional parallel trends based on never treated group. Third, they consider different ways to aggregate the average treatment effect on the treated estimators. Finally, in comparison with other available alternative to TWFE, the CS estimator is more flexible by allowing choosing the not yet treated as a control group and to add pre-treatment covariates (Baker et al., 2022).

Basically, CS estimator is a two-step estimation process and we keep the formalization as light as possible here but more details are available in the Appendix. The first step consists in estimating the unit cohort and time-specific treatment effects and the second step aggregating them to produce measures of overall treatment effects.

The first step is based on the following general formula of the ATT, with  $G$  the time period when a unit first becomes treated, for a unit  $g$  (group) at time  $t$ , the so-called group-time ATT:

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

Where  $\text{ATT}(g, t)$  is the expected difference between the observed outcome for the treated and for the untreated at time  $t$ . It allows for heterogeneity in ATT across cohorts or groups or over time.

The estimator estimates cohort time-specific treatment effects based on 2 by 2s comparison with adequate and “clean” controls. The treatment effect of a particular group can be estimated by the following regression using observation at time  $t$  and  $g-1$  from treated units and a set of clean controls:

$$Y^{g,t} = \alpha_1^{g,t} + \alpha_2^{g,t} \cdot G_g + \alpha_3^{g,t} \cdot 1\{T = t\} + \beta^{g,t} \cdot (G_g \times 1\{T = t\}) + \gamma \cdot X + \varepsilon^{g,t}$$

The group-time ATT is identified here following a conditional parallel trends assumption based on the never treated ( $nt$ ) group, estimate through the doubly robust estimator ( $dr$ ) (Sant’Anna & Zhao, 2020), here with a vector of pre-treatment covariates  $X$  measuring patient characteristics

(age quartile class, gender, NHI scheme, complementary public health insurance for the deprived population, and groups of diabetes drug therapy) and GP or patient list characteristics (age, gender, practice location 2SFA group, the number of registered patients encountered at least once during the year and their proportions: female, age 60-69 years old, age over 70, beneficiaries of the NHI health insurance for the deprived population, and beneficiaries of the salaried worker NHI schemes).

Also, it is worth noting that standard errors are clustered at the GP level.

The second step aggregates ATT based on the following general formula for aggregated ATT (AGGTT):

$$(2) \quad AGGTT(g, t) = \frac{\sum(w_{g,t} * ATT(g, t))}{\sum(w_{g,t})}$$

where  $w_{(g,t)}$  is a weight of how much information was used to estimate  $ATT(g, t)$ .

Finally, several robustness checks and heterogeneity analyses were implemented to confirm our results. First, we present the parallel trend tests proposed by Callaway and Sant'Anna. Second, by comparing the PP subsample results with the ITT subsample one, we expect to estimate if the results are driven by APNs involvement in specific follow-up procedures for the patient and/or by improvement of GPs practice pattern regarding T2DM management.

Third, we ran additional checks with stratification according to a set of key variables to control for heterogeneity in the effects of the pilot program. These key variables are the following:

- Quartiles of the above-mentioned LPA to consider the balance between the supply and demand at the GP living area level; and
- Clusters of GP-APN pairs using the taxonomy from Afrite *et al.* (Afrite et al., 2019), only for the subsample of 203 GPs who participated in this 2015 survey, to estimate the impact of the pilot while considering the heterogeneity in the nature and intensity of cooperation and skill mixing that this taxonomy revealed with a distinction between "firmly established" or "developing" pairs versus "under development pairs".

Unfortunately, we cannot implement stratified models according to socioeconomic status by distinguishing between the beneficiaries of NHI health insurance for the deprived population due to the small sample size and lack of power to ensure the model's convergence.

## **Results**

### **Description of the populations**

GPs' characteristics included age, sex, overbilling or not, the number of patients encountered at least once a year, and urban/rural area. Patients' characteristics included age group and percentage of patients with universal and/or complementary NHI schemes for the deprived population. Table 1 shows the 2010 characteristics of GPs who entered *Asalée* between 2012 and 2015 and their controls before and after matching. Before matching, the treated (N=612) and control groups (1561) differ significantly and the treated group exhibits notably a larger proportion of GPs located in rural and remote location and a larger number of encountered patients, compared to the control group. After matching, the treated (N=435) and control (N=973) groups were strictly comparable with a few exceptions regarding first the geographical locations of GPs' practices – with a higher percentage of treated GPs practicing in rural and remote areas – and second the GPs' T2DM patients' characteristics with a higher percentage of patients with long-term chronic illness and beneficiaries of NHI compulsory schemes for low-income individuals. Table 2 shows the 2010 characteristics of T2DM patients who entered *Asalée* between 2012 and 2015. The ITT samples before matching, with 18,638 treated and 43,011 control groups, were not quite comparable with many differences regarding age, NHI schemes, geographical location, or oral or insulin therapy for T2DM. After matching for both ITT and PP subsamples, the treated and control groups were strictly comparable, with one exception regarding a significantly higher mean number of visits with the GPs during the year for the T2DM patients of the control group (7.6 vs 7.1).

**Table 1: GPs' descriptive statistics before and after matching**

	Before coarsened Exact Matching		After coarsened Exact Matching	
	Control	Treated	Control	Treated
<b>GP's characteristics</b>				
Age (mean)	<b>51.6</b>	<b>51.1</b>	48.9	48.9
% male	<b>79.4</b>	<b>76.6</b>	66.2	65.7
<b>GP's Practice Location Areas</b>				
% Periurban	<b>21.8</b>	<b>20.1</b>	22.9	21.4
% Rural and remote	<b>30.8</b>	<b>43.8</b>	<b>25.7</b>	<b>42.1</b>
% Retirement and tourism or % Urban deprived	<b>23.7</b>	<b>8.9</b>	<b>22.6</b>	<b>7.1</b>
% Mixed inner cities or % Privileged cities/suburbs	<b>23.8</b>	<b>27.2</b>	28.8	29.4
% Localized potential accessibility - quartile 1	<b>21.8</b>	<b>20.1</b>	23.1	19.5
% Localized potential accessibility - quartile 2	<b>30.8</b>	<b>43.8</b>	<b>26.5</b>	<b>31.5</b>
% Localized potential accessibility - quartile 3	<b>23.7</b>	<b>8.8</b>	24.9	25.5
% Localized potential accessibility - quartile 4	<b>23.8</b>	<b>27.3</b>	25.4	23.5
<b>GP's Encountered Patient characteristics</b>				
Nb. Patient encountered at least once during the year (mean)	<b>1510.9</b>	<b>1650.6</b>	1607.3	1621.6
% female	56.2	56.1	55.4	55.4
% < 15 years old	18.9	18.7	23.1	22.7
% 16-59 years old	<b>47.2</b>	<b>46.0</b>	52.5	52.2
% 60-69 years old	<b>13.2</b>	<b>13.4</b>	9.9	10
% > 70 years old	<b>20.7</b>	<b>21.8</b>	14.1	14.6
% long-term chronic illness	<b>23.3</b>	<b>24.2</b>	<b>16.2</b>	<b>16.9</b>
% long-term chronic illness - diabetes	<b>2.4</b>	<b>2.7</b>	<b>3.5</b>	<b>3.9</b>
% NHI schemes for low income individuals	<b>3.6</b>	<b>4.0</b>	<b>3.5</b>	<b>4.0</b>
% free complementary insurance for low income individuals	7.5	7.5	7.4	8.0
% salaried workers NHI scheme	<b>75.2</b>	<b>74.1</b>	75	74.5
% self-employed worker NHI scheme	5	5.1	5.0	4.9
% agricultural worker NHI scheme	<b>8.5</b>	<b>8.8</b>	7.2	7.7
% other insurance scheme	<b>11.3</b>	<b>12</b>	12.7	12.9
<b>N obs</b>	<b>1561</b>	<b>612</b>	<b>973</b>	<b>435</b>

Note: Bold, mean or percentage, indicate statistically significant differences (F or T tests at .05) between treated and control groups before and after matching.

Table 2: T2DPs' descriptive statistics before and after matching

	Initial T2DPs samples <sup>a</sup>		Intention to treat subsamples <sup>a</sup>		Per protocol subsamples <sup>a</sup>	
	Before coarsened exact matching		After coarsened exact matching		After coarsened exact matching	
	Control <sup>b</sup>	Treated <sup>b</sup>	Control <sup>b</sup>	Treated <sup>b</sup>	Control <sup>b</sup>	Treated <sup>b</sup>
<b>Sociodemographic and insurance</b>						
Age (mean)	65.24	64.82	63.72	63.77	61.03	61.44
% age - quartile 1	33.93	33.39	33.40	33.40	37.27	37.27
% age - quartile 2	25.99	25.70	25.71	25.71	32.08	32.08
% age - quartile 3	23.23	23.23	23.23	23.23	22.32	22.32
% age - quartile 4	<b>16.85</b>	<b>17.67</b>	17.67	17.67	08.32	08.32
% male	53.34	54.14	54.19	54.19	53.82	53.82
<b>Health insurance</b>						
% NHI schemes for low-income individuals	03.81	03.79	03.30	03.30	02.92	02.92
% free complementary insurance for low-income individuals	<b>01.57</b>	<b>01.89</b>	01.53	01.53	01.46	01.46
% salaried worker compulsory NHI scheme	<b>76.93</b>	<b>75.67</b>	76.26	76.26	75.53	75.53
% farm worker compulsory NHI scheme	<b>09.87</b>	<b>10.48</b>	10.33	10.33	10.50	10.49
% self-employed worker compulsory NHI scheme	05.37	05.54	05.32	05.32	05.10	05.10
% other compulsory NHI scheme	<b>07.83</b>	<b>08.30</b>	08.09	08.09	08.87	08.87
<b>Living areas<sup>c</sup></b>						
% Periurban	<b>21.76</b>	<b>17.79</b>	17.91	17.91	19.61	19.61
% Rural and remote	<b>31.61</b>	<b>50.44</b>	50.32	50.32	49.51	49.51
% Retirement and tourism or % Urban deprived	<b>22.69</b>	<b>07.22</b>	07.26	07.26	07.20	07.20
% Mixed inner cities or % Privileged cities/suburbs	23.93	24.56	24.51	24.51	23.68	23.68

<b>Therapy</b>						
No. of visits with GPs (mean)	<b>7.62</b>	<b>7.13</b>	<b>7.26</b>	<b>7.04</b>	7.20	7.17
% Insulin therapy	<b>08.14</b>	<b>09.02</b>	08.79	08.79	06.18	06.18
% Oral anti-diabetics - first line therapy	24.11	24.32	24.49	24.49	24.60	24.60
% Oral anti-diabetics - second line dual therapy	15.48	15.99	15.98	15.98	18.01	18.01
% Oral anti-diabetics - third triple therapy	09.28	08.94	08.75	08.75	11.72	11.72
% Oral anti-diabetics - others	<b>01.52</b>	<b>01.27</b>	01.15	01.15	01.49	01.49
<b>N. Obs</b>	<b>43011</b>	<b>18638</b>	<b>18310</b>	<b>18310</b>	<b>2943</b>	<b>2943</b>

Note:

<sup>a</sup> See the flowchart for the details regarding the definitions of the samples.

<sup>b</sup> A boldface mean or percentage indicates statistically significant differences (F or t tests at .05) between treated and control groups.

<sup>c</sup> According to a typology of living areas.

Sources: *Asalée* not-for-profit organization registry; SNDS and SNIR-PS 2010-2017 (*Assurance Maladie*).

Figures of the trends based on the event study analysis (see Fig. 2) of the yearly proportion of patients correctly followed up for each of the process of care outcomes show that the average proportion of diabetes patients correctly followed up is higher or equivalent to the average for the treated group before the intervention.

### **Results from DiD staggered models**

Estimates showed that enrollment within the *Asalée* pilot program had a significant and dramatic impact, *ceteris paribus*, but only for the PP subsample (see Fig 2, Table 3 and Fig 3 to 7 in the Appendix for PP subsample estimates and Table 4 in the Appendix for ITT estimates). We observed that compared to the control group and the initial values (before enrollment), T2DPs within the *Asalée* pilot and PP, were overall significantly better followed up after their GPs being enrolled for all the six outcomes with an ATT comprise between 2.2 to 10.15 percentage point increase. That being said, depending on the outcome, we also observed that this average treatment effects on the treated (ATT) is highly both group or cohort, and time specific. Also, some pre-trends are significant and positive before treatment for microalbuminuria and to a less extent for EKG (and/or cardiologist) and funduscopy (and/or ophthalmologist visit).

Regarding three of the four outcomes for which the initial average of T2DM is particularly low (*e.g.* HbA1C, albuminuria, EKG and/or cardiologist visit), the significant and positive impact is observed for all the cohort with one exception, there is no significant impact for the cohort 2013 regarding EKG and/or cardiologist visit.

Estimates for HbA1c exhibit a clear pattern over groups and time, without any significant positive pre-trend before treatment, with an overall ATT that increase over time (from 11.22 percentage point increase in year 1 to 17.05 in year 5), an ATT larger for earlier cohorts compared to later ones, an ATT that increase the first two (cohort 2015), three (cohort 2013, 2014) or fourth (cohort 2012) years of exposure and stabilized or decreased after.

Estimates for microalbuminuria are contrasted, with a significant and positive pre-trend one year before treatment (that concerned mainly the 2013 cohort), an overall ATT that globally increased over time but with a sinusoidal shape (from 6.85 percentage point increase in year 1 to 13.80 in year 5), an ATT larger for the cohort 2015, that increased almost constantly for the

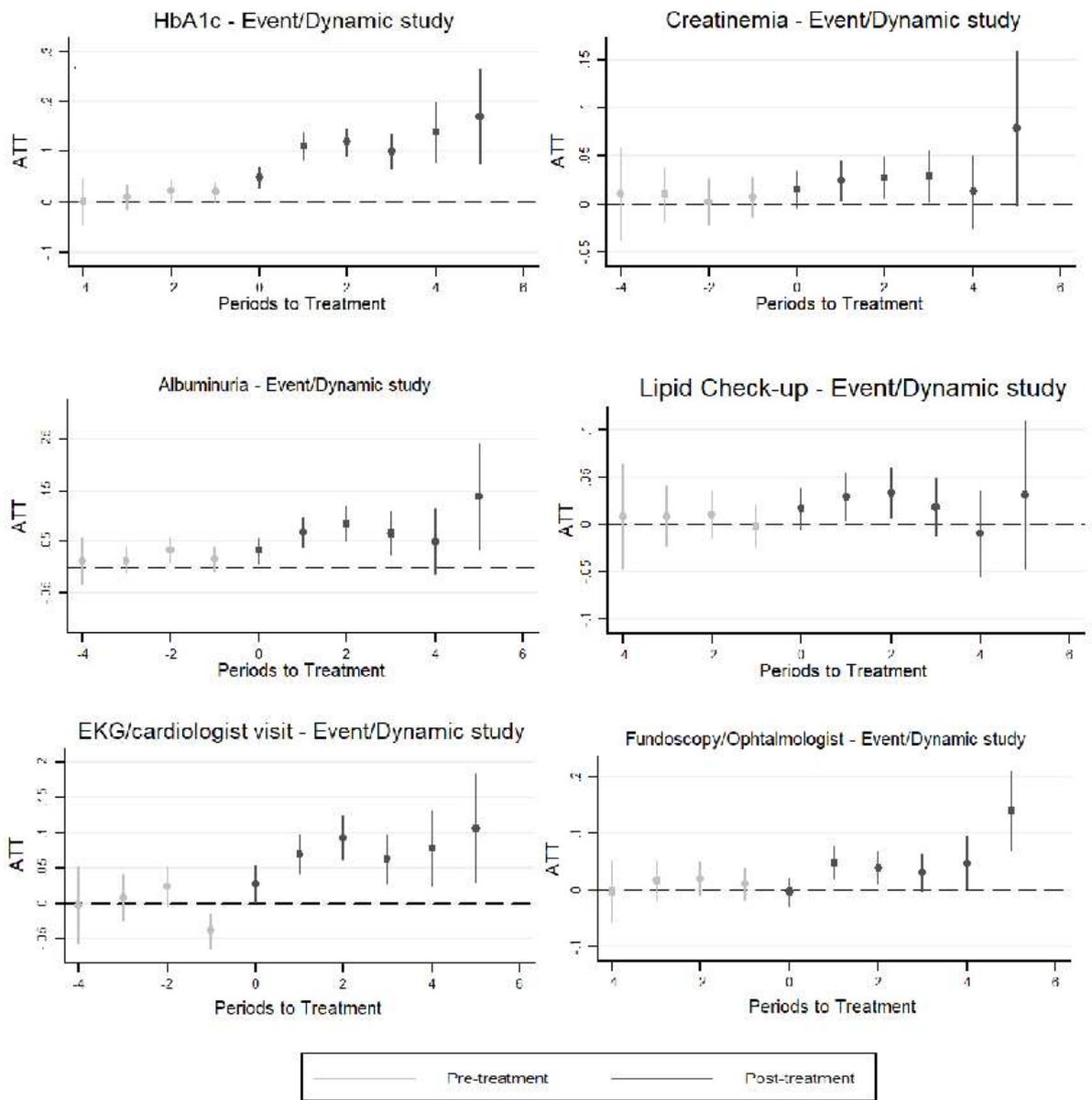
2012 cohort, only the first three years for cohorts 2013/2014 and then decreased a bit, increased the first year and stay constant after for the cohort 2015.

Estimates for EKG and/or cardiologist visit exhibit clear sinusoidal but increasing pattern, with no significant and positive pre-trends before treatment, an overall ATT that largely increased over time (from 6.96 percentage point increase in year 1 to 10.59 in year 5), quite larger and constant for the cohort 2012, less impressive but significant for the cohorts 2014 and 2015, non-significant for the cohort 2013.

Estimates for fundoscopy and/or ophthalmologist visit exhibit no clear patterns, with no significant and positive pre-trends before treatment, an overall ATT that globally increased over time especially after year 5 (around 4.8 percentage point increase in years 1 to 4, to 13.92 in year 5), but completely driven by the positive and significant result for the cohort 2012 and knowing that there is no significant and positive impact on ATT for the other cohorts.

Finally, for the two on six outcomes for which the initial average of T2DM is quite reasonable (e.g. creatinemia and lipid check-up), over 70 % of correctly followed patient, the significant and positive ATT is observed only for specific the cohorts 2014 and 2012 regarding creatinemia, 2014 and 2015 regarding lipid check-up.

Fig. 2 Per protocol subsample, ATT estimates for type two diabetes process indicators, event study over the period 2010-2017



We performed several robustness checks and heterogeneity analyses in addition to the parallel trend assumption tests. First, the results for ITT subsamples did not show any positive impact of the DiD variable, meaning that the estimated positive effect was due to enrollment in the *Asalée* pilot and the T2DM patients enrollment within a specific diabetic protocol program run by APNs following GPs referral. If the GPs, following the inclusion within *Asalée*, would have dramatically improved their T2DM management patterns for all T2DP, the impact would also have been observed for the ITT subsamples that is not the case here. Second, according to the heterogeneity analyses based on the PP subsamples and restricted to GPs participating in a 2015 survey, the impact of the pilot program is not different when considering the intensity and nature of the cooperation.

The results of stratified models regarding LPA that distinguish quartiles 1 to 3 (low- to medium-served living areas) versus 4 (“over”-served living areas) also show that the positive impact is not largely influenced by supply and demand balance.

## **Discussion**

Our main study purpose was to take a nationwide French pilot program of teams working between GPs and APNs over 2012-2015 to test whether supplementation and substitution improve the quality of care delivered for T2DPs.

Our study showed that, based on a quasi-experimental design with exact matching and difference-in-differences estimations, based on DiD staggered treatment design, GPs who entered the *Asalée* pilot program between 2012 and 2015 had significantly greater improvement in their T2DM patients regarding national process of care guidelines for the following: HbA1c, creatinemia, albuminuria, lipid check-up, funduscopy or a visit to an ophthalmologist, and EKG or a visit to a cardiologist. This positive and significant impact was only observed when a per protocol (versus intention to treat) perspective is considered and depending on the outcome, group and time. Then, one could conclude that the effects are mainly driven thanks to the involvement of the APNs for T2DP management in addition to the GPs and not only on GPs practice patterns improvement. Also, we estimate heterogeneity in the ATT. Specifically, when initial average of the outcome is particularly low (e.g. HbA1C, albuminuria, EKG and/or cardiologist visit), the significant and positive impact is larger and more systematic. On the contrary, when the initial average of T2DM is high and above 70 % of correctly followed patient (e.g. creatinemia and lipid check-up), the positive and significant impact depends on cohorts and time.

Several robustness checks confirmed these results, and stratification or subsample analysis that estimates the impacts depending on a small subsample of GPs (203) for which the nature and intensity of cooperation with APNs are known show equivalent positive impact of cooperation on the quality of care process for T2DPs.

Our study was the first to measure the impact of the *Asalée* pilot program on the quality of care delivered at the national level. Additionally, to the best of our knowledge, our study was the first study of such a scale (number of T2DM patients included) using such a robust quasi-experimental evaluation design and large panel data set linking claims and clinical data with a significant duration of follow-up. We used exhaustive NHI claims data with a hierarchical and panel structure that offers three advantages: facilitating the analysis of the change within units of observation, helping to establish the direction and magnitude of causal relationships, and improving the reliability and accuracy of measurements. Our study used a controlled before-after and quasi-experimental design combining coarsened exact matching – prior to the intervention at both the GP and T2DP levels – with difference-in-differences (DiD) estimates on a staggered treatment design to mitigate the possible risk of bias due to nonrandomization by considering the sample's initial differences in composition and unobserved heterogeneity constant over time. Our study offered a double perspective with intention to treat and per protocol T2DM patients' samples, reflecting the overall and specific change of GPs' practices regarding T2DP follow-up. Finally, our study focuses on process of care indicators since they are more sensitive to differences in the quality of care than outcome measures and easier to interpret. Finally, our results showed significant and dramatic effects of cooperation or skill mix and task shifting regarding the process of care indicators for which evidence in the literature was scarce (Maier et al., 2022b).

Our study also has several limitations. One limitation is the choice of the indicators to monitor the quality of health care. The debate over the most useful types of performance indicators is of international concern (Mant, 2001). It is simplistic to view process and outcome measures, indeed distal ones, as competing with each other, but there are circumstances where one type of measure is likely to be more useful than the other. In our case, the aim of our study was to assess the impact of this pilot program on GPs' professional practices and T2DP follow-up. Using process indicators in this case seemed particularly relevant because they directly represented what the GPs actually did (i.e., prescribe) when they were seeing patients. It is obvious that the dosage of glycosylated hemoglobin does not predict glycemic control. This represents that GPs address patients' diabetes during consultations and decides to act. This

also shows that the patient takes action on the doctor's prescription after the visit. These are behaviors of interest as far as the *Asalée* pilot is concerned. Finally, although we could have better discussed intermediate outcome indicators (HbA1c, microalbuminuria, creatinemia, lipid check-up, fundoscopy, and EKG), our study still reflected changes in biological status that could affect subsequent health outcomes, as demonstrated by Wolters and colleagues regarding HbA1c (Wolters et al., 2017).

Selection bias may also be a limitation since GPs voluntarily enrolled in the pilot. Then, our result could overestimate the positive impact of the pilot, thanks to the selection of GPs which would be motivated by joining the pilot due to interest in better monitoring T2DM patients and/or cooperating with APNs. If our quasi-experimental design is able to reduce bias, it does not solve all questions regarding the generalizability of our results since the effect of unobservable variables on the treated and control groups could potentially change over time. However, qualitative research on the *Asalée* pilot program has shown that if APN participation was not random, GP involvement did not systematically rely on specific motivations (Fournier et al., 2018; Afrite et al., 2019). The two consecutive extensions of the pilot at the regional and national levels were clearly less controlled by the *Asalée* NGO, and enrollment focus on the practice more than the physician level, which caused a massive and less restricted entry of GPs that ends with heterogeneous expectations of the GPs regarding the pilot. Also, the difference in results depending on the perspective, only positive for the PP subsample, support the interpretation that the effects are mainly driven by GPs-APN cooperation than by selection of GPs. That being said, it is fair to consider the fact that, first, APN selection could occur since APN were trained, hired and paid by the *Asalée* NGO, and second, that we cannot disentangle from the overall positive impact what is coming from recruitment, training, incentives and/or cooperation with GPs.

Despite these limitations, we can still draw some policy implications. First, teamwork between GPs and APNs in the *Asalée* pilot program context leads to a significant improvement in the quality of care delivered by GPs in terms of process of care indicators and this complement other positive effects of the pilot regarding GPs technical efficiency (Loussouarn et al., 2020). The improvement of the adequacy with guidelines regarding the process of care indicators calls for some comments. In 2018, the level of adequacy with guidelines for process of care indicators was much greater than those observed by the French national survey on diabetes patients in 2019 and above mention for all the indicators except for creatinemia and lipid check-up, which have already been relatively well followed up nationwide. The difference is particularly large and impressive for HbA1c. Second, we show that the impact is higher

when the initial level of adherence is low.

Considering disease prevention and taking a patient-centered care perspective raised the question of what other indicators could be more representative of the patients' real lives. The *Asalée* pilot program was expected to change decision-making processes and to make patients more involved in their care. However, our study does not allow us to measure such goals. This would require either shared decision-making or patient empowerment/autonomy to be integrated instead of (or in addition to) biological parameters. Additional patient-centered outcome indicators could be considered: satisfaction and quality of life. New indicators and new ways of collecting these indicators might be necessary. The Diabetes Quality of Life (DQOL) provides a total health-related quality of life score that examines the patient's direct experience regarding how well they are managing their diabetes and how they are feeling generally. It predicts self-reported satisfaction with diabetes control and care behaviors as effectively as possible (Burroughs et al., 2004). However, these indicators would be very useful, but there is currently no French national database capable of collecting these types of data.

Finally, expanding the scope of the pilot could generate more quality improvement and the decision of the Ministry of Health to promote and support advanced practice nurses' education and positions seems to move in this direction (Debout, 2018).

## References

- Abadie, A. (2005). Semiparametric difference-in-difference estimators. *The Review of Economic Studies* 72(1), 1–19.
- Afrite, A., Franc, C., Mousquès, J. (2019). *Des organisations et des pratiques coopératives diverses entre médecins généralistes et infirmières dans le dispositif Asalée: Une typologie des binômes*. Questions d'économie de la santé n°239, Paris.
- Baker, A. C., Larcker, D. F., Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144 (2), 370-395. <https://doi.org/10.1016/j.jfineco.2022.01.004>.
- Barlet, M., Coldefy, M., Collin, C., & Lucas-Gabrielli, V. (2012). *Local potential accessibility (LPA): A new measure of accessibility to private general practitioners*. IRDES, Questions d'économie de la Santé n°174, Paris.
- Barnett, R. (2018). Type 1 diabetes, *The Lancet*, 391(10117), 195. [http://dx.doi.org/10.1016/s0140-6736\(18\)30024-2](http://dx.doi.org/10.1016/s0140-6736(18)30024-2)
- Bertrand, M., Duflo, E., Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. <http://dx.doi.org/10.3386/w8841>
- Burroughs T.E., Desikan, R., Waterman, B.M., Gilin, D., McGill, J. (2004). Development and validation of the diabetes quality of life brief clinical inventory. *Diabetes Spectrum*, 17, 41-9. <http://dx.doi.org/10.2337/diaspect.17.1.41>
- Callaway, B., Sant'Anna, P.H.C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, Volume, 225(2), 200-230. <https://doi.org/10.1016/j.jeconom.2020.12.001>.
- Campbell, S.M. (2003). Improving the quality of health care: Research methods used in developing and applying quality indicators in primary care. *BMJ*; 326(7393), 816–9. <http://dx.doi.org/10.1136/bmj.326.7393.816>
- Carey, N., Courtenay, M., James, J., Hills, M., Roland, J. (2008). An evaluation of a Diabetes Specialist Nurse prescriber on the system of delivering medicines to patients with diabetes. *JCN*, 17, 1635-4. <https://doi.org/10.1111/j.1365-2702.2007.02197.x>
- Chaput, H., Monziols, M., Fressard, L., Verger, P., Ventelou, B., Zaytseva, A. (2019). *Plus de 80 % des médecins généralistes libéraux de moins de 50 ans exercent en groupe*. DREES, Etudes et résultats n°1114, Paris.
- Chevillard, G., Mousquès, J., Lucas-Gabrielli, V., Rican, S. (2019). Has the diffusion of primary care teams in France improved attraction and retention of general practitioners in rural areas? *Health Policy*, 123, 508–515. <https://doi.org/10.1016/j.healthpol.2019.03.002>
- Chevreur, K., Berg Brigham, K. & Bouché, C. (2014). The burden and treatment of diabetes in France. *Globalization and Health*, 10(1). <http://dx.doi.org/10.1186/1744-8603-10-6>

- Collins, S., Hewer, I. (2014). The impact of the Bologna process on nursing higher education in Europe: a review. *Int J Nurs Stud*, 51, 150-6. <https://doi.org/10.1016/j.ijnurstu.2013.07.005>
- de Bont, A., van Exel, J., Coretti, S., et al. (2016). Reconfiguring health workforce: a case-based comparative study explaining the increasingly diverse professional roles in Europe. *BMC Health Serv Res*, 8 (16), 637. <http://dx.doi.org/10.1186/s12913-016-1898-0>
- Debout, C. (2018). Infirmière de pratique avancée en France : première esquisse. *Soins*, 63(824):59–65. <http://dx.doi.org/10.1016/j.soin.2018.03.001>
- DREES (2020). *Les dépenses de santé en 2019 - Résultats des comptes de la santé*. DREES, Panorama de la Drees, Paris.
- Fosse-Edorh, S., Mandereau-Bruno, L. (2015). Suivi des examens recommandés dans la surveillance du diabète en France, en 2013. *Bull Epidemiol Hebd*;(34-35), 645-54. [http://www.invs.sante.fr/beh/2015/34-35/2015\\_34-35\\_5.html](http://www.invs.sante.fr/beh/2015/34-35/2015_34-35_5.html).
- Fournier, C., Bourgeois, I., Naiditch, M. (2018). *Action de santé libérale en équipe (Asalée) : un espace de transformation des pratiques en soins primaires*. IRDES, Question d'économie de la santé n°232, Paris.
- Freund, T., Everett, C., Griffiths, P., Hudon, C., Naccarella, L., Laurant, M. (2015) Skill mix, roles and remuneration in the primary care workforce: Who are the healthcare professionals in the primary care teams across the world? *Int J Nurs Stud* ;52:727-43. <http://dx.doi.org/10.1016/j.ijnurstu.2014.11.014>
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Ho, D.E., Imai, K., King, G., Stuart, E.A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, 15, 199–236. <https://doi.org/10.1093/pan/mpl013>
- Houweling, S.T., Kleefstra, N., van Hateren, K.J., Groenier, K.H., Meyboom-de Jong B., Bilo, H.J. (2011). Can diabetes management be safely transferred to practice nurses in a primary care setting? A randomised controlled trial. *JCN*;20, 1264-72. <https://doi.org/10.1111/j.1365-2702.2010.03562.x>
- Iacus, S.M., King, G., Porro, G. (2011). Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, 106, 345–361. <https://doi.org/10.1198/jasa.2011.tm09599>
- Imbens, G.W., Wooldridge, J.M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47, 5–86. <http://dx.doi.org/10.1920/wp.cem.2008.2408>
- Kernick, D., & Scott, A. (2002). Economic approaches to doctor/nurse skill mix: problems, pitfalls, and partial solutions. *The British journal of general practice*, 52(474), 42–46. <https://doi.org/10.1201/9781315385525-18>
- King, G., & Nielsen, R. (2019). Why propensity scores should not Be used for matching. *Political Analysis*, 27(4), 435–454. <https://doi.org/10.1017/pan.2019.11>

- Kringos, D., Boerma, W., Bourgueil, Y., et al. (2013): The strength of primary care in Europe: an international comparative study. *British Journal of General Practice*, 63(616), e742–e750. <http://dx.doi.org/10.3399/bjgp13x674422>
- Laurant, M., van der Biezen, M., Wijers, N., Watananirun, K., Kontopantelis, E., van Vught, A. (2018). Nurses as substitutes for doctors in primary care. *Cochrane Database of Systematic Reviews*;7:CD001271. <http://dx.doi.org/10.1002/14651858.CD001271.pub3>
- Legendre, B. (2020). *En 2018, les territoires sous-dotés en médecins généralistes concernent près de 6% de la population*. Etudes et résultats n°1144, DRESS, Paris.
- Lenz, E.R., Munding, M.O., Hopkins, S.C., Lin, S.X., Smolowitz, J.L. (2002). Diabetes Care Processes and Outcomes in Patients Treated by Nurse Practitioners or Physicians. *The Diabetes Educator*, 28(4):590–8. <http://dx.doi.org/10.1177/014572170202800413>
- Loussouarn, C., Franc, C., Videau, Y., Mousquès J. (2020). Can General Practitioners Be More Productive? The Impact of Teamwork and Cooperation with Nurses on GP Activities. *Health Economics*, 30(3), 680–98. <http://dx.doi.org/10.1002/hec.4214>
- Maier, C.B., Aiken, L.H. (2016). Task shifting from physicians to nurses in primary care in 39 countries: a cross-country comparative study. *Eur J Public Health* 26, 927-34. <http://dx.doi.org/10.1093/eurpub/ckw098>
- Maier, C.B., Aiken, L.H., Busse R. (2017). Nurses in advanced roles: policy levers to implementation. OECD Health Working Paper. Paris, OECD.
- Maier, C.B., Williams, G.A., Budde, H., Pflirter, L., Busse, R., Wismar, M. (2022a). Skill-mix for primary and chronic care: definitions, conceptual framework and relevance for policy and practice. In: Maier C.B., Kroezen M., Busse R., Wismar M. editors. Skill-mix innovation, effectiveness and implementation: improving primary and chronic care. WHO, European Observatory on Health Systems and Policies. Cambridge: Cambridge University Press; p. 1-38. doi.org/10.1017/9781009031929
- Maier, C.B., Budde, H., Pflirter, L., Kroezen, M. (2022b). Skill-mix changes: what evidence on patient outcomes and health systems?. In: Maier C.B., Kroezen M., Busse R., Wismar M. editors. Skill-mix innovation, effectiveness and implementation: improving primary and chronic care. WHO, European Observatory on Health Systems and Policies. Cambridge: Cambridge University Press; p. 39-69. doi.org/10.1017/9781009031929
- Mant, J. (2001). Process versus outcome indicators in the assessment of quality of health care. *International Journal for Quality in Health Care*, 13, 475-80. <https://doi.org/10.1093/intqhc/13.6.475>
- Martínez-González, N.A., Tandjung, R., Djalali, S., Rosemann, T. (2015). The impact of physician-nurse task shifting in primary care on the course of disease: a systematic review. *Hum Resour Health*, 7, 13-55. <http://dx.doi.org/10.1186/s12960-015-0049-8>
- Massin, S., Paraponaris A., Bernhard M., et al. (2014). *Les médecins généralistes face au paiement à la performance et à la coopération avec les infirmiers*. DREES, Etudes et Résultats n°873, Paris

- Mousquès, J., Bourgueil, Y., Le Fur, P., Yilmaz E. (2010). Effect of a French experiment of team work between general practitioners and nurses on efficacy and cost of type 2 diabetes patients care. *Health Policy*, 98, 131-143. <https://doi.org/10.1016/j.healthpol.2010.06.001>
- Nicholson, S., & Propper, C. (2011). Medical workforce. In Mark V. Pauly, Thomas G. McGuire & Pedro P. Barros (Eds.), *Handbook of health economics* (Vol. 2, pp. 873–925). Elsevier. <https://doi.org/10.1016/b978-0-444-53592-4.00014-1>
- OECD. (2015). *Cardiovascular Disease and Diabetes: Policies for Better Health and Quality of Care*. OECD Health Policy Studies, OECD Publishing, Paris. <https://dx.doi.org/10.1787/9789264233010-en>
- OECD. (2020). *Realising the Potential of Primary Health Care*. OECD Health Policy Studies, OCDE Publishing, Paris, <https://doi.org/10.1787/a92adec4-en>
- Parker, D., Maresco-Pennisi, D., Clifton, K., Shams, R., & Young, J. (2016). Practice nurse involvement in the management of adults with type 2 diabetes mellitus attending a general practice: results from a systematic review. *International journal of evidence-based healthcare*, 14(2), 41–52. <https://doi.org/10.1097/XEB.0000000000000080>
- Rabe-Hesketh, S., Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata Volume II: Categorical Responses, Counts, and Survival*. Third Edition. Stata Press, USA.
- Rios-Avila, F., Sant'Anna, P., Callaway, B. (2022). CSDID: Stata module for the estimation of Difference-in-Difference models with multiple time periods, <https://EconPapers.repec.org/RePEc:boc:bocode:s458976>.
- Rubin D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 688-701.
- Sant'Anna, P.H.C., Zhao, J. (2020). Doubly robust difference-in-differences estimators, *Journal of Econometrics*, 219 (1),101-122.<https://doi.org/10.1016/j.jeconom.2020.06.003>
- Sun, L. Abraham, S. (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*,225 (2), 175-199. <https://doi.org/10.1016/j.jeconom.2020.09.006>.
- Taylor, K.I., Oberle, K.M., Crutcher, R.A., Norton, P.G. (2005). Promoting Health in Type 2 Diabetes: Nurse-Physician Collaboration in Primary Care. *Biol Res Nurs*, 6, 207-15. <https://doi.org/10.1177/1099800404272223>
- Tsiachristas, A., Wallenburg, I., Bond, C.M., Elliot, R.F., Busse, R., van Exel J., et al. (2015). Costs and effects of new professional roles: Evidence from a literature review. *Health Policy*, 119(9):1176–87. <http://dx.doi.org/10.1016/j.healthpol.2015.04.001>
- Vrijhoef, H.J.M., Diederiks, J.P.M., Spreeuwenberg, C., Wolffenbuttel, B.H.R. (2001). Substitution model with central role for nurse specialist is justified in the care for stable type 2 diabetic outpatients. *J Adv Nurs*, 36, 546-55. <https://doi.org/10.1046/j.1365-2648.2001.02007.x>

Wolters, R.J., Braspenning, J.C.C., Wensing, M. (2017). Impact of primary care on hospital admission rates for diabetes patients: A systematic review. *Diabetes Research and Clinical Practice*, 129,182–96. <http://dx.doi.org/10.1016/j.diabres.2017.05.001>

## Appendix

### Callaway and Sant'Anna event study DiD estimator

CS estimator is a two-step estimation process. The first step consists in estimating the unit cohort and time-specific treatment effects and the second step aggregating them to produce measures of overall treatment effects.

The first step is based on the following general formula of the ATT, with  $G$  the time period when a unit first becomes treated, for a unit  $g$  (group) at time  $t$ , the so-called group-time ATT:

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

Where  $\text{ATT}(g, t)$  is the expected difference between the observed outcome for the treated and for the untreated at time  $t$ . It allows for heterogeneity in ATT across cohorts or groups or over time. Following the notation suggests by Rios-Avila<sup>2</sup>, with never treated group as counterfactual as it is the case here, the ATT, for the group ( $g$ ) at time  $t$ , could be specified as follows:

$$(2) \text{ ATT}(g, t) = [EY(g)_t - EY(NT)_t] - [EY(g)_{g-1} - EY(NT)_{g-1}]$$

where  $g$  is the index for treated cohorts or groups (the year when  $g$  is treated) and  $t$  is the year index ( $t=2010, \dots, 2017$ ),  $EY$  the expected value of the outcome, at time  $t$  or  $g-1$  (one year before the treatment), for the treated ( $g$ ) or never treated ( $NT$ ) groups. The first part calculates the differences in outcomes at time  $t$ , the second the difference at time  $g-1$ . When  $t < g$ ,  $\text{ATT}(g, t)$  can be used to test the robustness of the parallel trend assumption.

$$(3) m_{g,t}^{nt}(X) = \mathbb{E} [Y_t(0) - Y_{t-1}(0) | X, G_g = 1] = \mathbb{E} [Y_t(0) - Y_{t-1}(0) | X, NT = 1]$$

The group-time ATT is identified following a conditional parallel trends assumption based on the never treated ( $nt$ ) group (3), estimate through the doubly robust estimator ( $dr$ ) (Sant'Anna & Zhao, 2020), here with a vector of pre-treatment covariates  $X$  measuring patient characteristics (age quartile class, gender, NHI scheme, complementary public health insurance for the deprived population, and groups of diabetes drug therapy) and GP or patient list characteristics (age, gender, practice location 2SFA group, the number of registered patients encountered at least once during the year and their proportions: female, age 60-69 years old, age

---

<sup>2</sup> See Federico Rios-Avila website, [https://friosavila.github.io/playingwithstata/main\\_csdid.html](https://friosavila.github.io/playingwithstata/main_csdid.html)

over 70, beneficiaries of the NHI health insurance for the deprived population, and beneficiaries of the salaried worker NHI schemes).

Finally, under these above-mentioned assumptions, group time ATTs are semi-parametrically point-identified:

$$(4) \quad \overline{ATT}_{dr}^{nt}(g, t) = \mathbb{E} \left[ \left( \frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X) NT}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X) NT}{1-p_g(X)}\right]} \right) (Y_t - Y_{g-1} - m_{g,t}^{nt}(X)) \right]$$

This is a weighted (in the first part of the equation) average of the difference in the outcome. This re-weighting process aims to balanced covariates for both treated and control groups. To give the intuition of how it works let says that you up-weight or down-weight observations depending on how similar or not they are between treated and control groups, and you did this for each group  $g$  and time  $t$ .

Considering  $\beta^{g,t} = \overline{ATT}(g, t)$  under limited anticipation assumptions and homogenous treatment effects, it can be obtained by running the following population linear regression :

$$(5) \quad Y^{g,t} = \alpha_1^{g,t} + \alpha_2^{g,t} \cdot G_g + \alpha_3^{g,t} \cdot 1\{T = t\} + \beta^{g,t} \cdot (G_g \times 1\{T = t\}) + \gamma \cdot X + \varepsilon^{g,t}$$

Also, it is worth noting that standard errors are clustered at the GP level.

Since Callaway and Sant'Anna DiD specification for staggered treatment proposed numerous ATT's, they also propose useful aggregated ATT (AGGTT) that follows this general formula:

$$(6) \quad AGGTT(g, t) = \frac{\sum(w_{g,t} * ATT(g, t))}{\sum(w_{g,t})}$$

where  $w_{(g,t)}$  is a weight of how much information was used to estimate  $ATT(g, t)$ .

**Table 3: Average treatment effect on the treated (ATT) estimates for T2D PP patients, depending on the cohort and/or group**

Per Protocol Estimates <sup>a,b,c</sup>		Hba1c	Creatinemia	Microalbuminuria	Lipid check-up	EKG and/or cardiologist visit	Fundoscopy and/or ophthalmologist visit
Aggregated group-time ATT		0,1015 (0,0130) [0,000]	0,0243 (0,0096) [0,012]	0,0641 (0,0134) [0,000]	0,0221 (0,0107) [0,038]	0,0655 (0,0125) [0,000]	0,0336 (0,0112) [0,003]
Event Study							
	-4	0,0010 (0,0229) [0,966]	0,0107 (0,0247) [0,666]	0,0141 (0,0231) [0,542]	0,0089 (0,0285) [0,756]	-0,0023 (0,0280) [0,935]	-0,0022 (0,0273) [0,937]
	-3	0,0106 (0,0122) [0,383]	0,0099 (0,0148) [0,502]	0,0142 (0,0127) [0,263]	0,0088 (0,0162) [0,590]	0,0074 (0,0169) [0,660]	0,0170 (0,0182) [0,349]
	-2	0,0222 (0,0113) [0,049]	0,0023 (0,0122) [0,853]	0,0344 (0,0127) [0,007]	0,0108 (0,0123) [0,382]	0,0235 (0,0142) [0,098]	0,0203 (0,0148) [0,170]
	-1	0,0203 (0,0108) [0,060]	0,0075 (0,0106) [0,479]	0,0166 (0,0123) [0,175]	-0,0017 (0,0117) [0,885]	-0,0385 (0,0127) [0,002]	0,0105 (0,0145) [0,469]
	0	0,0497 (0,0110) [0,000]	0,0156 (0,0102) [0,126]	0,0331 (0,0132) [0,012]	0,0169 (0,0113) [0,136]	0,0272 (0,0137) [0,047]	-0,0032 (0,0127) [0,804]
	1	0,1122 (0,0135) [0,000]	0,0244 (0,0104) [0,019]	0,0685 (0,0147) [0,000]	0,0294 (0,0126) [0,020]	0,0696 (0,0141) [0,000]	0,0485 (0,0142) [0,001]
	2	0,1198 (0,0146) [0,000]	0,0275 (0,0112) [0,015]	0,0849 (0,0169) [0,000]	0,0333 (0,0134) [0,013]	0,0925 (0,0157) [0,000]	0,0402 (0,0140) [0,004]
	3	0,1020 (0,0177) [0,000]	0,0288 (0,0132) [0,030]	0,0667 (0,0221) [0,002]	0,0185 (0,0157) [0,238]	0,0627 (0,0178) [0,000]	0,0310 (0,0164) [0,059]
	4	0,1409 (0,0310) [0,000]	0,0129 (0,0195) [0,506]	0,0510 (0,0326) [0,118]	-0,0091 (0,0229) [0,691]	0,0783 (0,0274) [0,004]	0,0479 (0,0240) [0,046]
	5	0,1705 (0,0481) [0,000]	0,0788 (0,0411) [0,055]	0,1380 (0,0534) [0,010]	0,0320 (0,0398) [0,422]	0,1059 (0,0393) [0,007]	0,1392 (0,0360) [0,000]
Pre-trend test		[0,118]	[0,461]	[0,001]	[0,758]	[0,055]	[0,061]
Year time							
	T2012	0,0350 (0,0388) [0,367]	0,0629 (0,0324) [0,052]	0,0145 (0,0371) [0,697]	0,0009 (0,0371) [0,982]	0,1065 (0,0461) [0,021]	0,0121 (0,0411) [0,768]
	T2013	0,0753 (0,0206) [0,000]	-0,0141 (0,0183) [0,440]	0,0068 (0,0217) [0,755]	-0,0061 (0,0210) [0,771]	0,0555 (0,0257) [0,031]	0,0188 (0,0245) [0,443]
	T2014	0,0815 (0,0170) [0,000]	0,0254 (0,0134) [0,058]	0,0421 (0,0184) [0,022]	0,0155 (0,0153) [0,311]	0,0411 (0,0180) [0,023]	0,0459 (0,0177) [0,010]
	T2015	0,1111 (0,0153) [0,000]	0,0370 (0,0120) [0,002]	0,0749 (0,0166) [0,000]	0,0404 (0,0146) [0,006]	0,0632 (0,0158) [0,000]	0,0127 (0,0160) [0,430]
	T2016	0,1356 (0,0161) [0,000]	0,0308 (0,0125) [0,014]	0,0874 (0,0179) [0,000]	0,0335 (0,0145) [0,021]	0,0849 (0,0176) [0,000]	0,0498 (0,0152) [0,001]
	T2017	0,0903 (0,0166) [0,000]	0,0140 (0,0127) [0,271]	0,0736 (0,0206) [0,000]	0,0104 (0,0157) [0,505]	0,0660 (0,0168) [0,000]	0,0372 (0,0160) [0,020]
Group/Cohort specific							
	G-2012	0,1482 (0,0419) [0,000]	0,0497 (0,0281) [0,077]	0,0636 (0,0342) [0,063]	0,0369 (0,0286) [0,196]	0,1285 (0,0320) [0,000]	0,0903 (0,0328) [0,006]
	G-2013	0,1046 (0,0257) [0,000]	0,0102 (0,0175) [0,560]	0,0501 (0,0268) [0,061]	-0,0108 (0,0199) [0,586]	0,0375 (0,0259) [0,147]	0,0207 (0,0216) [0,338]
	G-2014	0,0843 (0,0166) [0,000]	0,0338 (0,0145) [0,020]	0,0597 (0,0208) [0,004]	0,0314 (0,0165) [0,058]	0,0687 (0,0181) [0,000]	0,0264 (0,0164) [0,108]
	G-2015	0,0888 (0,0208) [0,000]	0,0052 (0,0158) [0,743]	0,0986 (0,0239) [0,000]	0,0473 (0,0214) [0,027]	0,0495 (0,0231) [0,032]	0,0189 (0,0227) [0,405]

Table 3 (continued)  
Group & Time specific

G-2012						
t_2010_2011	0,0105 (0,0387) [0,786]	-0,0268 (0,0414) [0,518]	0,0559 (0,0390) [0,151]	-0,0065 (0,0383) [0,866]	-0,1228 (0,0476) [0,010]	-0,0331 (0,0530) [0,533]
t_2011_2012	0,0350 (0,0388) [0,367]	0,0629 (0,0324) [0,052]	0,0145 (0,0371) [0,697]	0,0009 (0,0371) [0,982]	0,1065 (0,0461) [0,021]	0,0121 (0,0411) [0,768]
t_2011_2013	0,1127 (0,0443) [0,011]	-0,0069 (0,0334) [0,838]	0,0116 (0,0384) [0,762]	0,0232 (0,0422) [0,582]	0,1691 (0,0405) [0,000]	0,1142 (0,0473) [0,016]
t_2011_2014	0,1501 (0,0517) [0,004]	0,0530 (0,0340) [0,119]	0,0434 (0,0512) [0,396]	0,0595 (0,0389) [0,126]	0,1688 (0,0455) [0,000]	0,0996 (0,0417) [0,017]
t_2011_2015	0,1887 (0,0480) [0,000]	0,0470 (0,0322) [0,144]	0,0654 (0,0509) [0,199]	0,0601 (0,0374) [0,108]	0,0903 (0,0419) [0,031]	0,0809 (0,0483) [0,094]
t_2011_2016	0,2321 (0,0531) [0,000]	0,0633 (0,0343) [0,065]	0,1084 (0,0440) [0,014]	0,0460 (0,0363) [0,204]	0,1304 (0,0427) [0,002]	0,0955 (0,0394) [0,015]
t_2011_2017	0,1705 (0,0481) [0,000]	0,0788 (0,0411) [0,055]	0,1380 (0,0534) [0,010]	0,0320 (0,0398) [0,422]	0,1059 (0,0393) [0,007]	0,1392 (0,0360) [0,000]
G-2013						
t_2010_2011	0,0021 (0,0230) [0,926]	0,0211 (0,0276) [0,445]	0,0403 (0,0224) [0,073]	0,0088 (0,0252) [0,728]	0,0161 (0,0261) [0,537]	-0,0049 (0,0275) [0,858]
t_2011_2012	0,0344 (0,0212) [0,105]	0,0374 (0,0214) [0,080]	0,0580 (0,0196) [0,003]	0,0279 (0,0221) [0,208]	0,0082 (0,0239) [0,730]	0,0617 (0,0264) [0,020]
t_2012_2013	0,0588 (0,0214) [0,006]	-0,0173 (0,0215) [0,422]	0,0047 (0,0258) [0,857]	-0,0190 (0,0233) [0,416]	0,0057 (0,0300) [0,850]	-0,0230 (0,0252) [0,361]
t_2012_2014	0,1133 (0,0287) [0,000]	0,0181 (0,0203) [0,373]	0,0781 (0,0324) [0,016]	0,0084 (0,0242) [0,727]	0,0257 (0,0301) [0,394]	0,0705 (0,0301) [0,019]
t_2012_2015	0,1353 (0,0299) [0,000]	0,0362 (0,0212) [0,087]	0,0887 (0,0334) [0,008]	0,0075 (0,0249) [0,763]	0,0543 (0,0306) [0,076]	-0,0076 (0,0285) [0,789]
t_2012_2016	0,1149 (0,0326) [0,000]	0,0231 (0,0233) [0,321]	0,0534 (0,0392) [0,173]	-0,0178 (0,0252) [0,479]	0,0469 (0,0322) [0,145]	0,0367 (0,0277) [0,186]
t_2012_2017	0,1005 (0,0350) [0,004]	-0,0093 (0,0227) [0,682]	0,0256 (0,0414) [0,536]	-0,0334 (0,0279) [0,230]	0,0553 (0,0339) [0,103]	0,0269 (0,0289) [0,352]
G-2014						
t_2010_2011	-0,0017 (0,0157) [0,915]	0,0277 (0,0198) [0,161]	0,0165 (0,0171) [0,335]	0,0248 (0,0228) [0,276]	-0,0111 (0,0222) [0,616]	0,0206 (0,0250) [0,411]
t_2011_2012	0,0312 (0,0169) [0,064]	-0,0065 (0,0174) [0,709]	0,0433 (0,0204) [0,034]	0,0008 (0,0213) [0,969]	0,0550 (0,0239) [0,021]	0,0334 (0,0249) [0,180]
t_2012_2013	0,0154 (0,0185) [0,405]	-0,0114 (0,0181) [0,528]	-0,0151 (0,0197) [0,445]	-0,0077 (0,0216) [0,720]	-0,0592 (0,0215) [0,006]	0,0066 (0,0247) [0,790]
t_2013_2014	0,0400 (0,0187) [0,032]	0,0222 (0,0181) [0,219]	0,0175 (0,0222) [0,430]	0,0074 (0,0198) [0,709]	0,0140 (0,0221) [0,527]	0,0137 (0,0218) [0,530]
t_2013_2015	0,1025 (0,0213) [0,000]	0,0465 (0,0173) [0,007]	0,0543 (0,0236) [0,022]	0,0331 (0,0217) [0,127]	0,0794 (0,0214) [0,000]	0,0241 (0,0220) [0,273]
t_2013_2016	0,1268 (0,0211) [0,000]	0,0392 (0,0176) [0,026]	0,0912 (0,0267) [0,001]	0,0544 (0,0208) [0,009]	0,1161 (0,0260) [0,000]	0,0553 (0,0209) [0,008]
t_2013_2017	0,0678 (0,0202) [0,001]	0,0272 (0,0173) [0,116]	0,0760 (0,0303) [0,012]	0,0306 (0,0223) [0,170]	0,0652 (0,0241) [0,007]	0,0124 (0,0220) [0,572]
G-2015						
t_2010_2011	0,0010 (0,0229) [0,966]	0,0107 (0,0247) [0,666]	0,0141 (0,0231) [0,542]	0,0089 (0,0285) [0,756]	-0,0023 (0,0280) [0,935]	-0,0022 (0,0273) [0,937]
t_2011_2012	0,0304 (0,0228) [0,182]	-0,0186 (0,0249) [0,456]	0,0106 (0,0210) [0,614]	-0,0169 (0,0267) [0,526]	0,0371 (0,0301) [0,217]	0,0113 (0,0307) [0,714]
t_2012_2013	0,0292 (0,0242) [0,228]	-0,0040 (0,0249) [0,873]	0,0140 (0,0271) [0,607]	0,0288 (0,0245) [0,240]	-0,0188 (0,0294) [0,522]	0,0265 (0,0320) [0,407]
t_2013_2014	0,0176 (0,0230) [0,446]	0,0217 (0,0203) [0,284]	0,0045 (0,0314) [0,887]	-0,0215 (0,0259) [0,406]	-0,0158 (0,0262) [0,547]	-0,0176 (0,0307) [0,566]
t_2014_2015	0,0622 (0,0233) [0,008]	0,0179 (0,0197) [0,365]	0,0975 (0,0282) [0,001]	0,0781 (0,0249) [0,002]	0,0343 (0,0292) [0,241]	-0,0160 (0,0298) [0,592]
t_2014_2016	0,1264 (0,0243) [0,000]	0,0103 (0,0183) [0,572]	0,1080 (0,0280) [0,000]	0,0489 (0,0247) [0,047]	0,0543 (0,0294) [0,064]	0,0332 (0,0295) [0,259]
t_2014_2017	0,0776 (0,0267) [0,004]	-0,0126 (0,0206) [0,540]	0,0904 (0,0335) [0,007]	0,0149 (0,0288) [0,605]	0,0600 (0,0283) [0,034]	0,0395 (0,0287) [0,169]

<sup>a</sup> Sample size: see flowchart for sample selection details

<sup>b</sup> Note : this table shows average treatment effects on the treated (ATTs), aggregated (overall) or depending on specific group and/or year time, for each specific outcome variable. Estimates are based on doubly robust estimation. Never treated units are always the control groups. H0 Pre-trend test correspond to : all pre-treatment are equal to 0. The rows: "group-specific", summarize ATT by group (or cohort) of first year of treatment (i.e. 2012, 2013, 2014, 2015), "t\_year1\_year2" indicate respectively the pre and post treatment periods ; "Event study", summarize ATT by the length of exposure; "Calendar time", summarize ATT by year ". T2DPs' covariates: age group (ref.: quartile 4 ), male, compulsory health insurance scheme for salaried workers (ref.: other schemes), free complementary health insurance coverage (ref: none), drug therapy (ref.: insulin) ; GPs' covariates: age group (ref.: quartile 3 and 4), male, total number of enrolled patients, % of patients: female, 60-69 y.o., over 70 y.o., compulsory health insurance scheme for salaried workers, free complementary health insurance coverage, quartile 1-3 of the Localized Potential Accessibility. Even if covariates are time-varying, only the base-period values are used for the estimation. Standard errors are robust to heteroskedasticity (clustered at the GPs level) and shown in parentheses, p-value for significance shown between square brackets [].

Source : Asalée not-for-profit organisation registry; SNDS and SNIR-PS 2010-2017 (Assurance Maladie)

Fig. 3 Per protocol subsample, ATT estimates for HbA1C, depending on the group or cohort and time, over the period 2010-2017

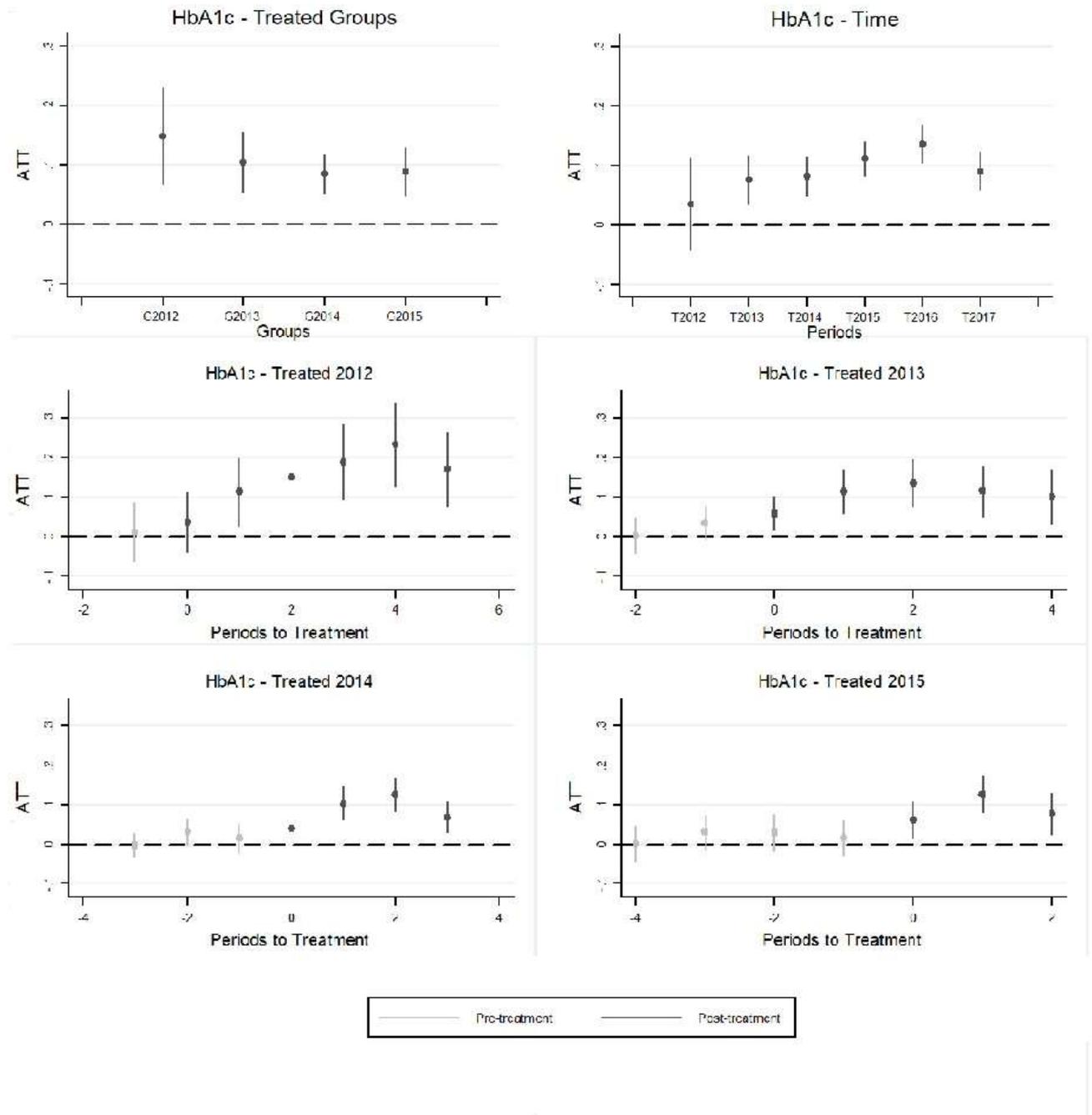


Fig. 4 Per protocol subsample, ATT estimates for creatinemia, depending on the group or cohort and time, over the period 2010-2017

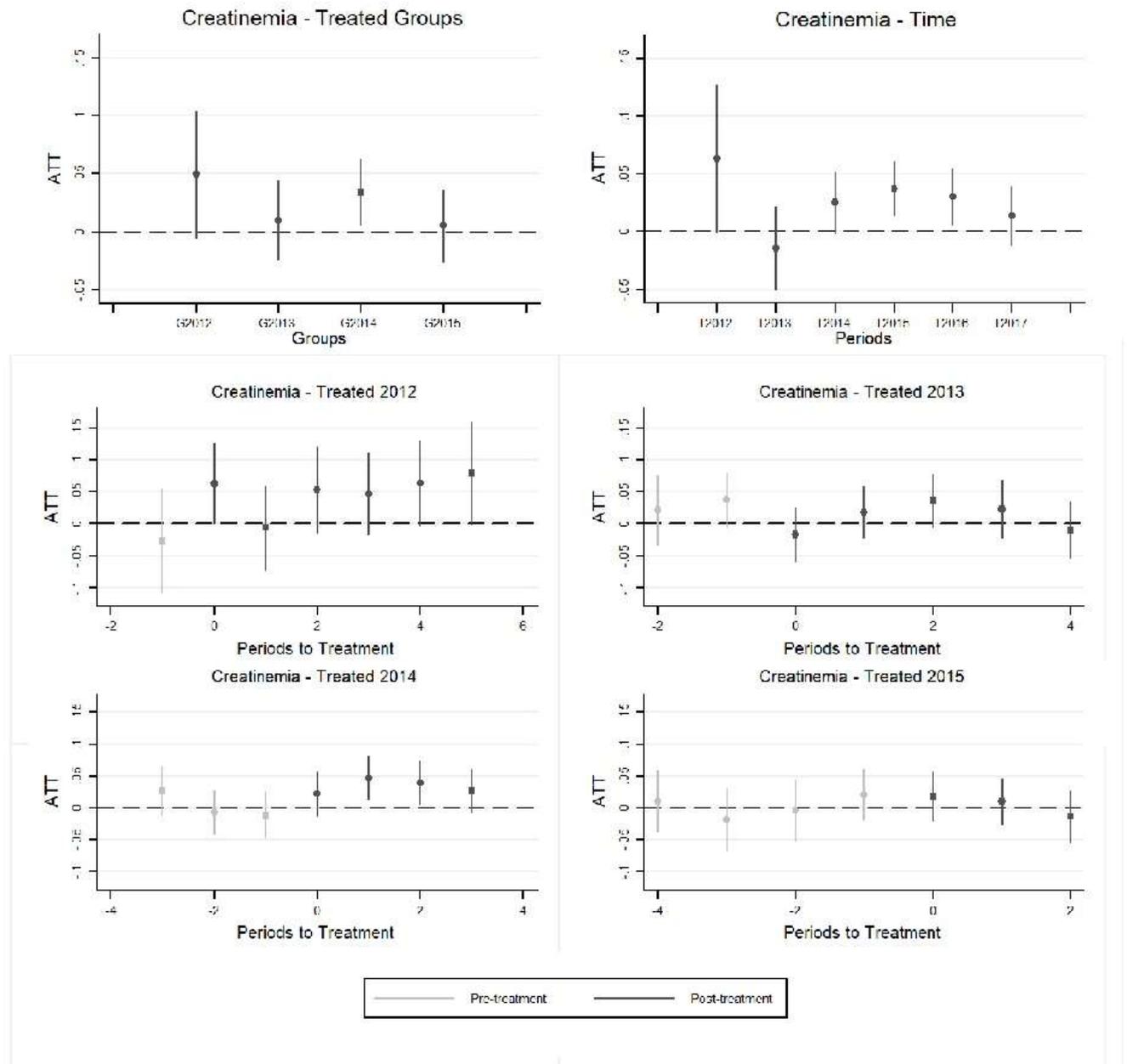


Fig. 5 Per protocol subsample, ATT estimates for albuminuria, depending on the group or cohort and time, over the period 2010-2017

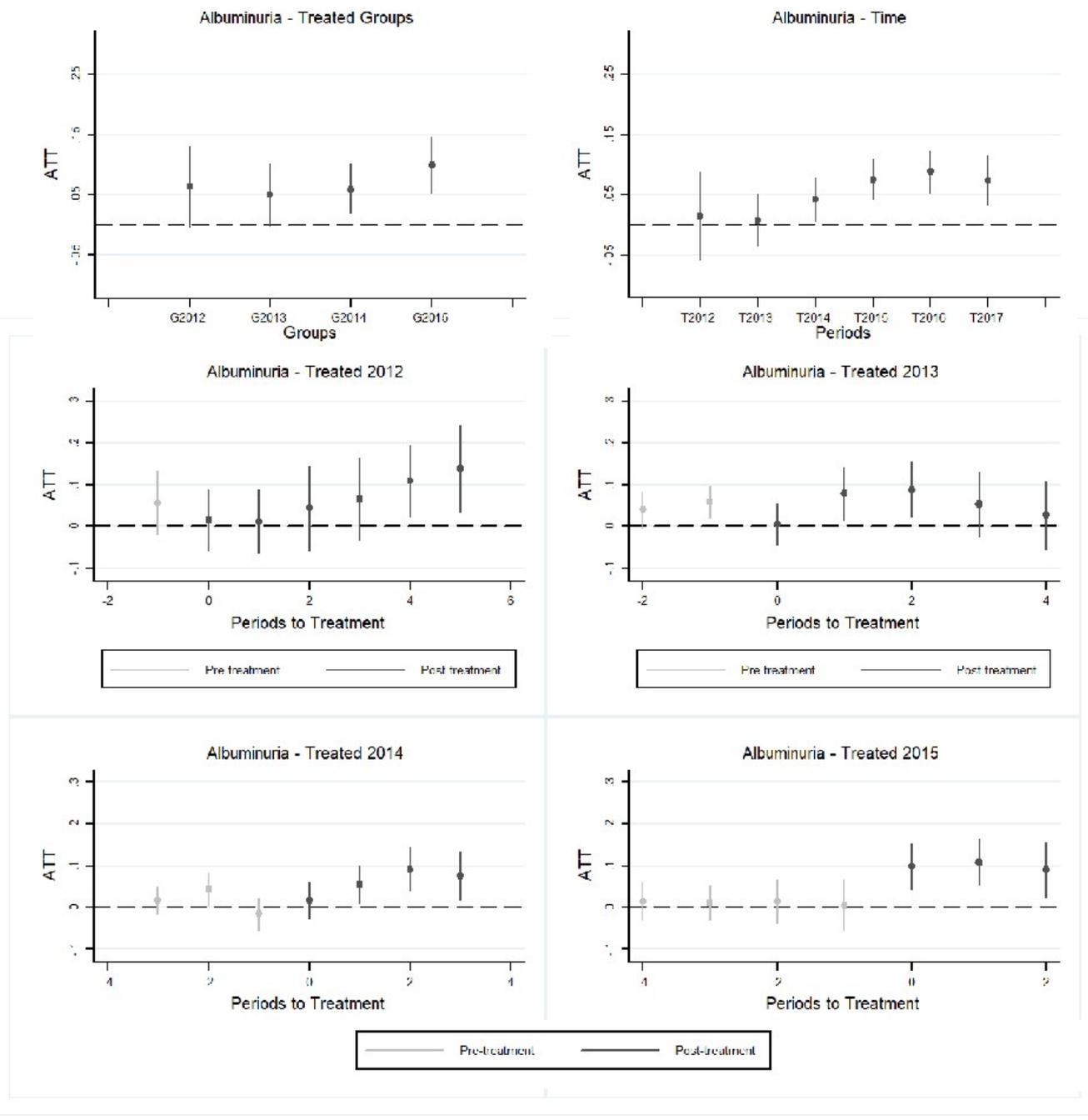


Fig. 6 Per protocol subsample, ATT estimates for lipid check-up, depending on the group or cohort and time, over the period 2010-2017

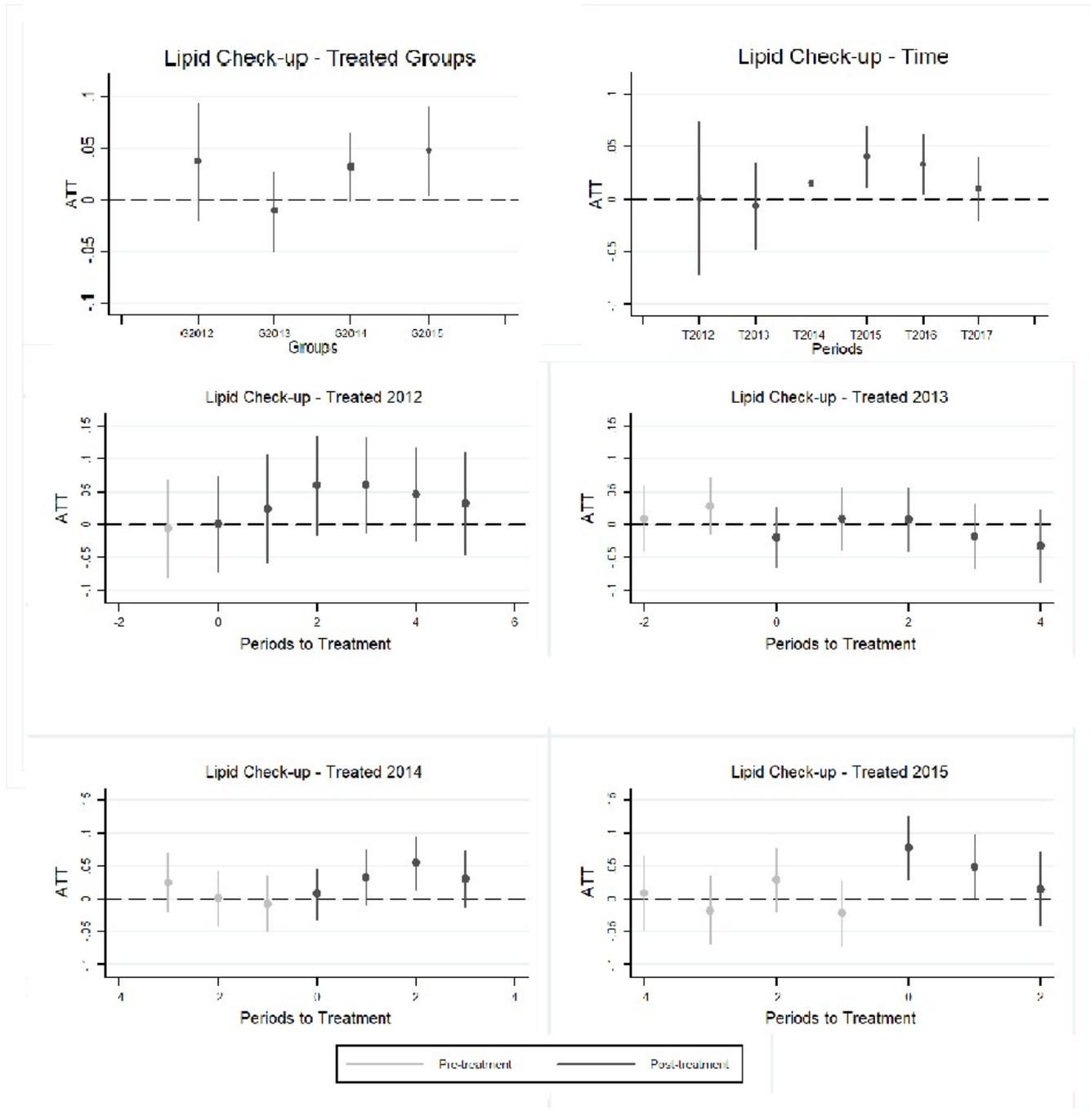


Fig. 7 Per protocol subsample, ATT estimates for EKG/cardiologist visit, depending on the group or cohort and time, over the period 2010-2017

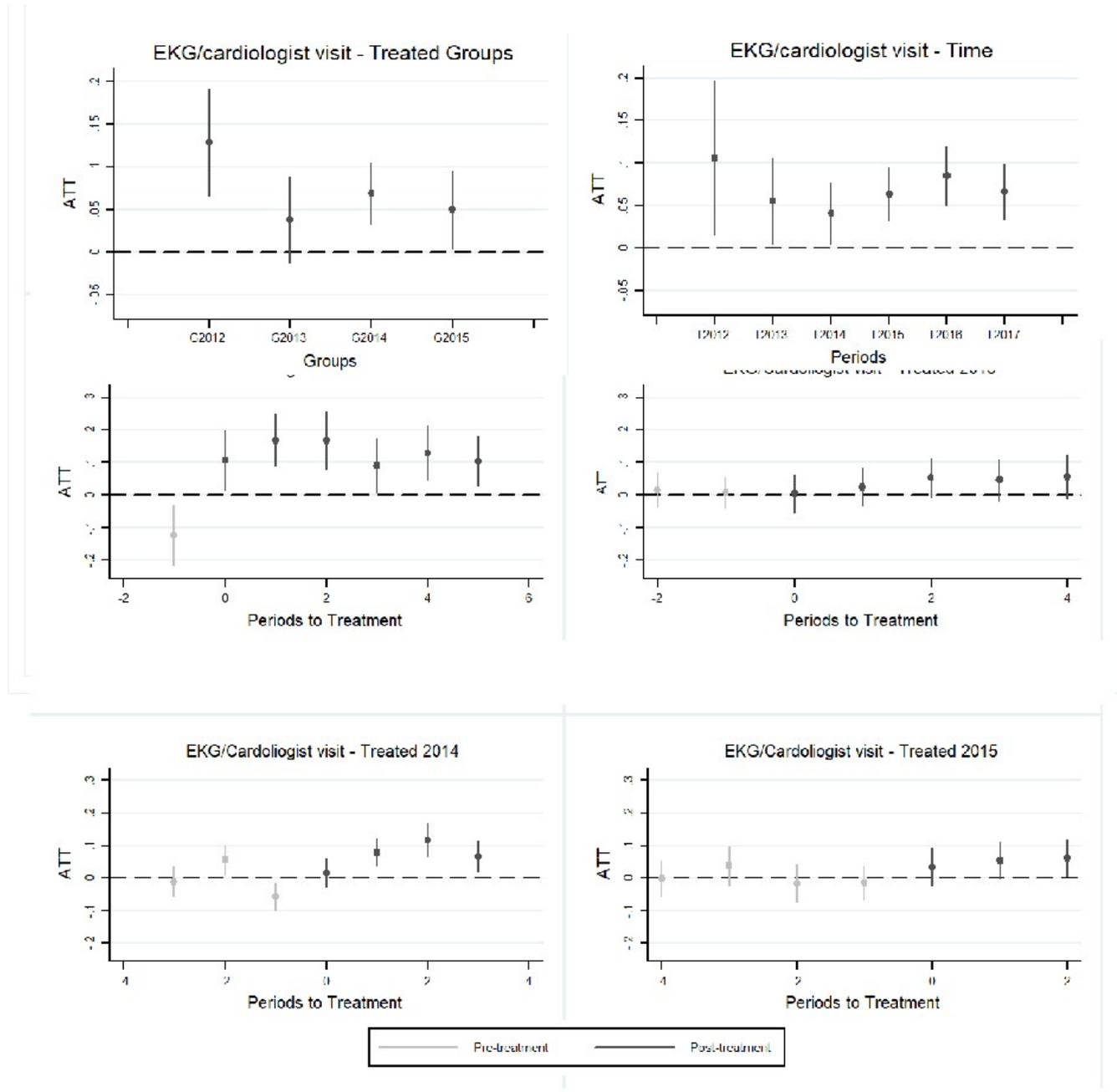
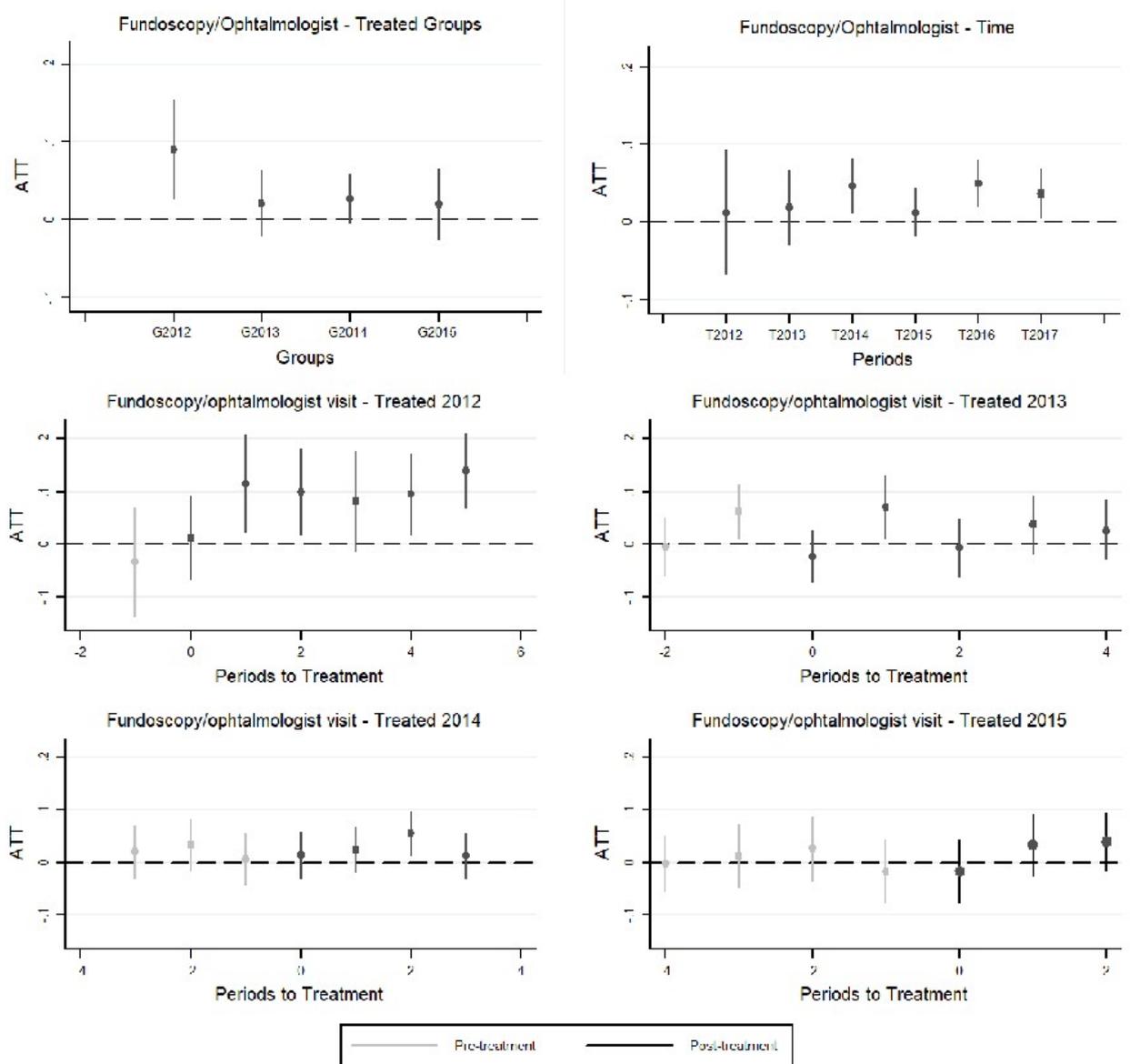


Fig. 8 Per protocol subsample, ATT estimates for funduscopy/ophthalmologist visit, depending on the group or cohort and time, over the period 2010-2017



**Table 4: Average treatment effect on the treated (ATT) estimates for T2D ITT patients**

Intention To Treat Estimates <sup>a,b,c</sup>	Hba1c	Creatinemia	Microalbuminuria	Lipid check-up	EKG and/or cardiologist visit	Fundoscopy and/or ophthalmologist visit
Aggregated group-time ATT	0,0071 -(0,0056) [0,205]	0,0096 (0,0057) [0,097]	0,0082 (0,0073) [0,260]	-0,0028 (0,0059) [0,639]	0,0180 (0,0051) [0,000]	0,0008 (0,0047) [0,857]
Pre-trend test	[0,796]	[0,317]	[0,620]	[0,404]	[0,113]	[0,181]

<sup>a</sup> Sample size: see flowchart for sample selection details

<sup>b</sup> Note : this table shows average treatment effects on the treated (ATTs), aggregated (overall) or depending on specific group and/or year time, for each specific outcome variable. Estimates are based on doubly robust estimation. Never treated units are always the control groups. H0 Pre-trend test correspond to : all pre-treatment are equal to 0. The rows: "group-specific", summarize ATT by group (or cohort) of first year of treatment (i.e. 2012, 2013,2014, 2015)," t\_year1\_year2" indicate respectively the pre and post treatment periods ; "Event study", summarize ATT by the length of exposure; "Calendar time", summarize ATT by year ". T2DPs' covariates: age group (ref.: quartile 4 ), male, compulsory health insurance scheme for salaried workers (ref.: other schemes), free complementary health insurance coverage (ref: none), drug therapy (ref.: insulin) ; GPs' covariates: age group (ref.: quartile 3 and 4), male, total number of enrolled patients, % of patients: female, 60-69 y.o., over 70 y.o., compulsory health insurance scheme for salaried workers, free complementary health insurance coverage, quartile 1-3 of the Localized Potential Accessibility. Even if covariates are time-varying, only the base-period values are used for the estimation. Standard errors are robust to heteroskedasticity (clustered at the GPs level) and shown in parentheses, p-value for significance shown between square brackets [].

Source : Asalée not-for-profit organisation registry; SNDS and SNIR-PS 2010-2017 (Assurance Maladie)