Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Adolescent extracurricular sport participation patterns and later healthcare use and expenditure: an 11-year linked cohort study in Hungary

Ágoston Koch

koch.agoston@tf.hu

Hungarian University of Sports Science

Tamás Sterbenz

Hungarian University of Sports Science

Erika Kendelényi-Gulyás

Hungarian University of Sports Science

Zoltán Bakucs

ELTE Centre for Economic and Regional Studies

Gergely Csurilla

Hungarian University of Sports Science

Research Article

Keywords: adolescence, extracurricular sport, healthcare utilization, prescriptions, healthcare expenditure, record linkage, longitudinal cohort

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11217733/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Evidence on adolescent sport and subsequent healthcare use remains fragmented across individual service sectors. We examined whether extracurricular sport participation patterns across two adolescent assessments were associated with service-specific healthcare use and expenditure through young adulthood.

Methods

We linked school assessments with population-panel and national health-insurance records to follow 23,283 Hungarian students for 11 years (2011–2021). Regular extracurricular sport was assessed in grades 8 (2008) and 10 (2010) and classified as reported at neither assessment, grade 10 only, grade 8 only, or both assessments. Poisson pseudo-maximum likelihood models estimated associations with annual utilization and expenditure, adjusting for age, sex, family background, and baseline-school and calendar-year fixed effects. False-discovery-rate adjustment covered the 39 main comparisons. Sensitivity analyses examined measured health and socioeconomic selection and alternative specifications.

Results

The analyses included more than 250,000 person-years, with 99.3% of the complete-case cohort observed in all 11 follow-up years. Compared with students reporting sport at neither assessment, those reporting sport at both had 17% fewer dispensed prescriptions (adjusted rate ratio 0.830, 95% confidence interval 0.770–0.895), fewer primary-care contacts (0.958, 0.934–0.983), and more specialist outpatient visits (1.074, 1.040–1.110). Patient pharmaceutical spending was lower (mean ratio 0.920, 0.882–0.961), but observable publicly reimbursed expenditure showed no clear association (1.000, 0.854–1.170). Grade-8-only participation was also associated with fewer prescriptions (0.880, 0.819–0.945). Prescription estimates changed little after adjustment for inter-wave healthcare use and pre-exposure parental histories; a separate grade-10 cohort showed a smaller association in the same direction.

Conclusions

Reporting sport at both adolescent assessments was associated with lower prescription and primary-care use alongside higher specialist outpatient use through young adulthood. These contrasting associations highlight the importance of examining service-specific healthcare use alongside aggregate expenditure when evaluating adolescent sport.

Background

Adolescence is a period of marked change in physical activity. Most young people do not meet current recommendations, and participation falls as they progress through secondary school [1, 2]. Regular sport is one of the principal organized settings in which adolescents are active, but it is not interchangeable with total physical activity. It combines repeated participation with access to facilities, adult supervision, and a social group, while entry and continuation depend on school provision, local opportunities, and household resources. A recent meta-analysis found generally favorable associations between youth sport and later physical activity, health, and wellbeing, but also emphasized the limited evidence from repeated exposure measures and the importance of selection into sport [3].

Sport status is not fixed during adolescence. Dropout is common [4], and tracking from youth sport or physical activity into adulthood is moderate rather than complete [5–8]. A single baseline measure therefore combines young people with very different histories: some continue, some stop, and others enter later [9]. Recent longitudinal studies have begun to model these patterns directly, showing that activity profiles across adolescence predict behavior and body composition in emerging adulthood [6, 7]. For subsequent healthcare use, repeated measurement is valuable not because two observations establish a continuous trajectory, but because they distinguish observed patterns that a one-time measure cannot.

Participation is also socially patterned. In high-income countries, children and adolescents from more advantaged households are substantially more likely to take part in organized sport [10, 11]. The same family resources, school environments, and health-related preferences can influence later access to care and medication use. Repeated sport measurement alone therefore does not resolve selection. A convincing analysis needs both a clear temporal ordering of sport and outcomes and information on the family circumstances that precede participation.

Healthcare use is a different outcome from health status. It reflects morbidity, access, clinical practice, and decisions to seek care, and the direction of association with sport may differ across services. Better fitness, psychosocial health, and related health behaviors may reduce primary-care or medication use. Sport may also lead to injuries, musculoskeletal complaints, diagnostic work-up, and specialist follow-up. Existing studies in children and adolescents have usually examined one part of this process: physician contacts [12], injury-related hospital care [13], psychotropic medication [14], or psychiatric hospitalization [15]. They do not show whether lower use in one sector coexists with higher use in another within the same population.

Evidence on expenditure is thinner and comes mainly from middle-aged or older adults. Physical-activity histories have been associated with lower later healthcare and pharmaceutical costs in Australian and US cohorts [16, 17], and with primary-care costs in a recent 10-year study [18]. These findings cannot be assumed to apply to young adulthood, when chronic disease remains relatively uncommon and sport-related contacts may offset lower use elsewhere. Nor can a single aggregate cost measure show which parts of the health system account for a difference.

Linked administrative records make a different longitudinal design possible. Stable identifiers preserve follow-up after participants leave school, claims files record service use and paid amounts independently of recall, and links to family and labor-market registers can recover pre-exposure household histories [19–21]. Because such records reflect registration, financing, and coding processes rather than a research protocol,

valid inference also requires an explicit denominator and transparent reporting of linkage and sample selection [22–24]. Hungary's Linked Administrative Panel Database (Admin4) follows a representative 50% sample of the population from 2003 to 2021 across education, employment, family, and healthcare modules [25, 26]. Linked to repeated school sport reports, it allows several service sectors to be followed in parallel, distinguishes a year with no recorded use from a year outside observation, and provides family histories measured before the first sport assessment.

This design addresses several limitations of the existing literature within a single linked cohort. We combine sport assessments in grades 8 and 10 with 11 years of administrative healthcare follow-up for 23,283 young people, distinguishing four observed participation patterns during adolescence. The analyses draw on more than 250,000 person-years, with 99.3% of the complete-case cohort observed in all 11 follow-up years. This breadth and continuity of observation allow us to examine primary care, specialist outpatient care, inpatient care, prescriptions, and expenditure in parallel within the same individuals through young adulthood. We then examine whether the estimates persist after accounting for healthcare use between the two sport assessments and for parental socioeconomic and employment histories measured before either outcome follow-up or the second sport assessment. Finally, we assess the grade-10 association in a separate, older school cohort with longer follow-up. The resulting question is deliberately narrower than whether sport is beneficial in general: are adolescent sport-participation patterns associated with the subsequent composition of healthcare use, and do any service-specific differences translate into lower observable publicly reimbursed expenditure? This matters for public health: if service sectors move in different directions, the healthcare implications of adolescent sport cannot be inferred from prescriptions or aggregate expenditure alone.

Methods

Data

We conducted a longitudinal record-linkage cohort study using the National Assessment of Basic Competencies (Országos Kompetenciamérés; OKM) and the Linked Administrative Panel Database (Admin4). Admin4 is constructed from records supplied by Hungarian administrative authorities and follows a representative 50% sample of the population from 2003 through 2021 [25]. The sampling frame originates from the national social-security register and is updated for entries to and exits from the population. Its core panel contains demographic and labor-market records; supplementary modules cover education, family relationships, social transfers, and healthcare. The data provider links the source registers before researcher access and retains a stable anonymized person identifier across years and modules. Analyses were conducted in the secure ELTE Centre for Economic and Regional Studies Databank environment; the research team had no access to direct identifiers. The representativeness of the Admin4 sampling frame does not imply that the final OKM-linked complete-case cohort is nationally representative, so each stage of cohort construction is reported explicitly.

The source cohort comprised 53,495 students observed in grade 8 in the 2008 OKM. We located the same anonymized individuals in the 2010 grade-10 assessment. Of the source cohort, 45,800 were observed in

grade 10, 25,923 had a valid broad sport measure at both assessments, and 23,286 also had complete information on age, sex, the family background index, and baseline school. Follow-up began in 2011, after the second sport assessment, and ended in 2021. A total of 23,283 students contributed at least one observed follow-up year. Main analyses included a person-year only when the individual was present in the Admin4 population panel in that year. Outcome-specific samples differed slightly because the PPML estimator removes fixed-effect singletons for sparse outcomes. Figure 1 documents the cohort flow, and Additional file 1: Table S2 reports the stage-specific counts and selection diagnostics. The study size was determined by the number of source-cohort members meeting the prespecified linkage, exposure, covariate, and follow-up criteria; no a priori sample-size calculation was performed.

At each assessment, students reported whether they regularly participated in (i) a school-organized extracurricular recreational sport activity and (ii) sport lessons or activities outside school. Neither item referred to compulsory physical education [27]. An explicit yes to either item defined sport participation; classification as no sport required explicit no responses to both. Undocumented code 0, codes 7–9, and system-missing values were treated as missing. We classified students as reporting sport at neither assessment, grade 10 only, grade 8 only, or both assessments. The categories describe status at two observations; they do not identify the date of entry or dropout or establish uninterrupted participation between waves. A stricter measure requiring valid responses to both sport items at both assessments was used in sensitivity analysis. Detailed exposure coding and the broad and setting-specific participation-pattern distributions are provided in Additional file 1: Table S1.

Annual healthcare measures came from National Health Insurance Fund modules covering primary care, specialist outpatient care, inpatient care, and dispensed prescriptions. Utilization outcomes were primary-care contacts, all specialist outpatient visits, laboratory services, non-laboratory outpatient visits, inpatient days, and prescriptions dispensed. Expenditure outcomes were outpatient, laboratory, non-laboratory outpatient, and inpatient expenditure; public pharmaceutical subsidy; patient pharmaceutical spending; and observable publicly reimbursed expenditure. The last measure summed outpatient expenditure, inpatient expenditure, and public pharmaceutical subsidy. It excludes person-level primary-care financing, fully private care, and costs outside the available modules. The H1 modules contain annual aggregate counts and expenditure but no diagnosis, procedure, or drug-class codes.

The healthcare event modules are sparse: person-months without a recorded event are absent from the event file [26]. After annual aggregation, we therefore coded a missing healthcare value as zero only when the individual was present in the Admin4 population panel in that year. Years without population-panel observation remained missing. This rule preserves the extensive margin of use while avoiding the classification of an unobserved year as zero utilization. Monetary values were converted to constant 2021 Hungarian forints using the national consumer price index [28]. Rare negative annual amounts, reflecting retrospective administrative corrections, were omitted only from the corresponding PPML expenditure outcome.

The main covariates were age at the grade-8 assessment, sex, and the continuous OKM family background index, which summarizes parental education and household resources. Baseline-school fixed effects absorb all time-invariant characteristics shared by students from the same school, including local sport provision

and school context; calendar-year fixed effects absorb common changes in healthcare use, reimbursement, coding, and cohort age. Parent-child links derived from family-allowance records were not required for inclusion in the main cohort. Because both the child and parent had to be present in the underlying 50% sample, at least one parent was linked for approximately half of the analytic cohort. Within that subsample, pre-exposure variables summarized parental earnings position, earnings-day intensity, labor-market activity, linked-parent composition, and reconstructed family links during 2003–2007. Reporting follows STROBE and RECORD [23, 29]; completed checklists are provided in Additional files 2 and 3, respectively. Additional file 3 also provides supplementary information on study-population selection, person-level linkage, data cleaning, and access to analytic materials.

Empirical strategy

For each outcome *k*, we estimated a Poisson pseudo-maximum likelihood (PPML) model on the annual person-year panel. Let *i* index students, *s* their grade-8 school, and *t* calendar year. The conditional mean was:

$$E\left[Y\square\square\square\square \mid S\square,\ X\square,\ \alpha\square,\ \lambda\square\right] = exp(\beta\square\square S\square\square\square\square + \beta\square\square S\square\square\square + \beta\square\square S\square\square\square\square\square + X\square\prime\gamma\square + \alpha\square + \lambda\square)$$

1

Sport reported at neither assessment is the omitted category. $S\square^{10}_i$, $S\square^{8}_i$, and $S\square\square\square\square_i$ indicate sport reported at grade 10 only, grade 8 only, or both assessments. X_i contains age, sex, and the family background index; $\alpha\square$ and $\lambda\square$ denote baseline-school and calendar-year fixed effects. Exponentiated sport coefficients are conditional rate ratios for utilization outcomes and ratios of expected annual expenditure for monetary outcomes. Because the sport-pattern indicators are fixed over follow-up, an individual fixed-effect model would absorb the exposure and was not used.

Identification comes from differences in sport pattern between students from the same baseline school, after accounting for common calendar-year shocks and observed covariates. School fixed effects remove time-invariant school-origin differences but not unmeasured differences between students within a school. The coefficients are therefore adjusted prospective associations, not treatment effects. In particular, the model cannot eliminate confounding by prior health, motivation, family health practices, peer networks, or access to sport that is not captured by the observed data.

PPML was used for both utilization and expenditure because it models the conditional mean on the outcome's original scale, retains zero values, and remains consistent under general forms of heteroskedasticity when the conditional mean is correctly specified [30, 31]. It does not require a Poisson variance. This avoids the zero-handling and retransformation problems of log-linear models and targets the annual mean, which is the relevant quantity for both service use and expenditure. Standard errors were two-way clustered by person and baseline school to allow serial correlation within students and residual dependence among students from the same school.

We estimated separate models for 13 outcomes and retained all three sport contrasts. Dispensed prescriptions and observable publicly reimbursed expenditure were designated as the focal utilization and

expenditure outcomes. Benjamini-Hochberg false-discovery-rate q values were calculated across the 39 main sport-outcome comparisons [32]. For the two focal outcomes, we also estimated stepwise specifications that added age and sex and then family background in the main sample, followed by parental-history adjustment in the core parent-linked sample. Because baseline-school and calendar-year fixed effects were integral to the design, both sets of fixed effects were retained in the least-adjusted specification.

Secondary and sensitivity analyses (dup: abstract ?)

Secondary analyses examined heterogeneity by sex and follow-up period and separated school-organized from out-of-school sport. Grade-specific models used all participants with a valid sport measure at the relevant grade, irrespective of validity at the other assessment. Period-specific models covered 2011–2013, 2014–2016, 2017–2019, and 2020–2021. We also repeated the grade-10 analysis in a separate cohort of students observed in grade 10 in 2008 and followed from 2009 through 2021.

Sensitivity analyses changed one design element at a time: the stricter two-item sport definition; exclusion of 2020–2021; baseline-school-by-year or sex-by-year fixed effects; restriction to students observed in all 11 follow-up years; stabilized inverse-probability weights for inclusion in the two-assessment cohort, trimmed at the first and 99th percentiles; and alternative clustering schemes. In the parent-linked subsample, we first repeated the main model without parental covariates and then added the 2003–2007 parental histories. A separate model added the child's 2009 primary-care contacts, non-laboratory outpatient use, inpatient days, and prescriptions, transformed with the inverse hyperbolic sine. Because 2009 lies between the two sport assessments, this last specification addresses inter-wave health selection rather than confounding measured before the first exposure. Analyses were conducted in Stata 17 using ppmlhdfe. Detailed specifications and complete results for the secondary and sensitivity analyses are provided in Additional file 1.

Results

Study population and sport patterns

Of the 53,495 students in the source cohort, 85.6% were observed again in grade 10. Among the 25,923 students with valid sport status at both assessments, 37.2% reported sport at neither assessment, 9.7% at grade 10 only, 24.7% at grade 8 only, and 28.4% at both (Additional file 1: Table S1). Students in the both-assessments group were more often male and had a higher family background index than those in the neither-assessment group. The complete-case cohort contained 23,286 students, of whom 23,283 contributed follow-up. The outcome models included 22,827–23,283 participants and 250,539–255,545 person-years; 99.3% of the complete-case cohort was observed in all 11 outcome years (Figure 1; Table 1; Additional file 1: Tables S2 and S3).

Table 1. Characteristics of the final analytic cohort by sport-participation pattern.

Characteristic	Neither	Grade 10 only	Grade 8 only	Both	SMD: G10	SMD: G8	SMD: both
Participants, N	8,615	2,280	5,706	6,682			
Age at grade 8 assessment, mean (SD)	13.84 (0.60)	13.85 (0.50)	13.85 (0.69)	13.85 (0.49)	0.02	0.02	0.02
Male, percent	37.6	44.9	47.1	55.6	0.15	0.19	0.37
Family background index, mean (SD)	-0.02 (0.90)	0.22 (0.93)	-0.00 (0.92)	0.31 (0.94)	0.26	0.02	0.35
At least one linked parent, percent	51.3	48.9	50.9	51.0	-0.05	-0.01	-0.01
Primary care contacts in 2009, mean (SD)	5.99 (5.31)	5.52 (4.56)	6.05 (5.00)	5.47 (4.67)	-0.09	0.01	-0.10
Non-laboratory outpatient visits in 2009, mean (SD)	2.84 (5.50)	2.97 (5.31)	2.85 (5.36)	3.37 (6.29)	0.02	0.00	0.09
Inpatient days in 2009, mean (SD)	0.31 (2.02)	0.27 (1.86)	0.31 (2.22)	0.28 (1.63)	-0.02	-0.00	-0.02
Dispensed prescriptions in 2009, mean (SD)	5.37 (33.63)	4.93 (24.56)	4.71 (8.58)	4.61 (9.00)	-0.01	-0.03	-0.03

Note. *The final analytic cohort has valid sport status at both assessments, complete baseline covariates, and at least one observed follow-up year.*

SMD denotes the standardized mean difference relative to participants reporting no sport at either assessment.

Healthcare use in 2009 occurred after the grade 8 assessment and before the grade 10 assessment.

The 2009 measures are descriptive and enter only the inter-wave health-selection sensitivity analysis.

Main findings

The clearest difference concerned prescriptions (Table 2; Figure 2). Relative to students reporting sport at neither assessment, those reporting sport at both filled 17% fewer prescriptions during follow-up (adjusted rate ratio 0.830, 95% confidence interval 0.770–0.895). They also had fewer primary-care contacts, but more specialist outpatient visits, particularly non-laboratory outpatient care. These associations remained among the findings that passed false-discovery-rate adjustment (Additional file 1: Table S4). Laboratory use and inpatient days did not differ clearly.

The service-specific differences were not accompanied by a clear difference in public expenditure (Table 3; Figure 2). Patient pharmaceutical spending was lower in the both-assessments group (mean ratio 0.920, 95% confidence interval 0.882–0.961), whereas no clear association with observable publicly reimbursed

expenditure was detected (1.000, 0.854–1.170); the interval is compatible with both lower and higher expenditure. Estimates for public pharmaceutical subsidy and inpatient expenditure were imprecise (Additional file 1: Table S4).

Lower prescription use was also observed among students reporting sport in grade 8 only (rate ratio 0.880, 0.819–0.945), while the grade-10-only estimate was close to the null. For prescriptions, the both-assessments group also had lower expected use than the grade-10-only group, whereas the direct comparison with the grade-8-only group remained uncertain (Additional file 1: Table S5).

Table 2. Adjusted associations between adolescent sport-participation patterns and annual healthcare utilization

Outcome	**Mean per person-year**	**Sport at grade 10 only**	**Sport at grade 8 only**	**Sport at both assessments**
		RR (95% CI)	RR (95% CI)	RR (95% CI)
Primary-care contacts	4.04	0.972 (0.939–1.007)	0.993 (0.968–1.019)	0.958 (0.934–0.983)†
Specialist outpatient visits, all	3.96	1.051 (1.003–1.102)	1.030 (0.997–1.065)	1.074 (1.040–1.110)†
Laboratory services	0.91	1.024 (0.960–1.093)	1.006 (0.962–1.052)	1.014 (0.969–1.061)
Non-laboratory outpatient visits	3.05	1.059 (1.010–1.110)	1.038 (1.002–1.075)	1.093 (1.056–1.130)†
Inpatient days	0.42	0.868 (0.706–1.066)	1.062 (0.899–1.254)	0.935 (0.789–1.109)
Dispensed prescriptions	4.01	0.949 (0.864–1.043)	0.880 (0.819–0.945)†	0.830 (0.770–0.895)†

Note. RR, adjusted rate ratio; CI, confidence interval. Reference group: no regular extracurricular sport reported at either assessment (grade 8 in 2008, grade 10 in 2010); outcomes cover 2011–2021.

Estimates are exponentiated PPML coefficients adjusted for age at the grade-8 assessment, sex, family background index, baseline-school fixed effects, and calendar-year fixed effects; 95% CIs are based on standard errors two-way clustered by person and baseline school.

Missing claims-module aggregates were coded as zero only in person-years with population-panel observation. Outcome-specific models included 22,827–23,283 participants and 250,539–255,545 person-years (Additional file 1: Table S3).

† Benjamini–Hochberg q < 0.05 across all 39 main sport–outcome comparisons; q values are reported in Additional file 1: Table S4.

Table 3. Adjusted associations between adolescent sport-participation patterns and annual healthcare expenditure

Outcome	Mean per person-year, HUF	Sport at grade 10 only	Sport at grade 8 only	Sport at both assessments
		MR (95% CI)	MR (95% CI)	MR (95% CI)
Outpatient expenditure	9,314	1.088 (1.014–1.167)	1.010 (0.969–1.052)	1.040 (0.998–1.083)
Laboratory expenditure	2,071	1.126 (0.955–1.328)	0.983 (0.918–1.053)	0.997 (0.924–1.076)
Non-laboratory outpatient expenditure	7,245	1.076 (1.012–1.143)	1.018 (0.977–1.061)	1.052 (1.010–1.097)
Inpatient expenditure	13,746	1.009 (0.749–1.359)	1.031 (0.922–1.153)	1.046 (0.925–1.184)
Public pharmaceutical subsidy	8,251	1.163 (0.727–1.862)	0.723 (0.389–1.346)	0.890 (0.528–1.503)
Patient pharmaceutical spending	3,462	0.985 (0.927–1.047)	0.967 (0.924–1.011)	0.920 (0.882–0.961)†
Observable publicly reimbursed expenditure	31,041	1.077 (0.877–1.323)	0.934 (0.780–1.119)	1.000 (0.854–1.170)

Note. MR, ratio of expected annual expenditure (mean ratio); CI, confidence interval. Reference group, covariates, fixed effects, clustering, and † are as in Table 2. Expenditures are in constant 2021 HUF.

Observable publicly reimbursed expenditure equals outpatient expenditure plus inpatient expenditure plus public pharmaceutical subsidy; person-level primary-care financing and fully private care are not observed. Rare negative annual administrative corrections were omitted only from the corresponding outcome. Outcome-specific models included 22,827–23,283 participants and 250,539–255,544 person-years (Additional file 1: Table S3).

Secondary and sensitivity analyses

Stepwise models indicated that adjustment for age and sex attenuated the prescription differences, with little further change after adding family background; observable publicly reimbursed expenditure estimates moved toward the null after demographic adjustment and remained imprecise (Additional file 1: Table S6). The prescription association was otherwise highly stable. Its magnitude changed little after adding pre-exposure parental socioeconomic and employment histories (Additional file 1: Table S7), the child's 2009 healthcare use (Additional file 1: Table S8), or under alternative exposure, follow-up, fixed-effect, sample, weighting, and clustering specifications (Additional file 1: Tables S9 and S10).

There was no evidence that sport-pattern associations differed by sex for either focal outcome (Additional file 1: Table S11). Formal interaction tests indicated variation across follow-up periods for prescriptions ($p = 0.0473$) and observable publicly reimbursed expenditure ($p = 0.0223$). The inverse prescription association for the both-assessments group was present in each period and was greater in magnitude in later periods, whereas the period-specific public-expenditure estimates were imprecise and showed no consistent temporal pattern (Figure 3; Additional file 1: Table S12). Grade-specific models indicated lower prescription use for any sport reported at either assessment, while setting-specific analyses showed lower estimates particularly for sport reported at both assessments; public-expenditure estimates remained imprecise (Additional file 1: Table S13). In the independent cohort of 34,112 students observed in grade 10 in 2008, regular extracurricular sport was associated with a smaller prescription difference in the same direction, while observable publicly reimbursed expenditure again showed no clear association (Additional file 1: Table S14).

Discussion

Regular extracurricular sport reported across grades 8 and 10 was associated with a distinct pattern of healthcare use through young adulthood. Students reporting sport at both assessments filled fewer prescriptions and used less primary care, yet had more specialist outpatient contacts. No clear association with observable publicly reimbursed expenditure was detected. The prescription association was reproduced across alternative exposure definitions and samples, remained after adjustment for inter-wave child healthcare and pre-exposure parental histories, and appeared in a separate school cohort.

The prescription finding is the clearest point of contact with earlier register studies. Adolescent activity has been associated with later antidepressant and hypnotic use [14] and with psychiatric hospitalization [15]. Our outcome covers all dispensed prescriptions, so it reflects a broader component of treated morbidity, but the available data cannot identify which therapeutic groups account for the difference. The result should therefore not be read as evidence about mental-health medication, analgesics, respiratory treatment, or any other class. It shows a durable difference in aggregate prescription use that warrants examination with drug-level data.

The contrasting associations across healthcare sectors are a central finding. Lower prescription and primary-care use coexisted with higher specialist outpatient use within the same cohort. A single healthcare endpoint would therefore give an incomplete account of the association between adolescent sport and subsequent healthcare use. Injury-related assessment is one plausible explanation, consistent with linked evidence that active children have more injury-related hospital contact [13], but diagnosis and procedure codes were unavailable. Rehabilitation, diagnostic investigation, access, and care-seeking may also contribute. The linked design shows that these sectoral differences occur within the same cohort; it cannot determine the clinical pathway behind them.

The family-linked analyses address an important source of concern in youth-sport research. Sport was socially patterned, yet restricting the sample to linked families and adding parental earnings, employment, and family-link histories measured before 2008 barely changed the prescription estimates. Nor did adjustment for the child's healthcare use between the two sport assessments. These results make

measured socioeconomic selection and inter-wave health selection less plausible as complete explanations. They do not rule out confounding by baseline health, parental preferences, peer networks, local sport supply, or other unobserved characteristics.

The linked data determine both the reach and the interpretation of the study. Sport was reported in school before follow-up; healthcare outcomes were generated independently in the insurance system; and the population panel distinguished a source-observed year with no recorded care from a year outside observation. Family linkage supplied histories measured before the first sport assessment. Together, these features reduce reliance on recall and conventional cohort retention and make it possible to compare several parts of the health system within the same individuals. They do not make the records purpose-built research data: the outcomes remain shaped by registration, reimbursement, and coding, and relevant behavioral and clinical detail is absent.

The lack of a clear association with observable public expenditure contrasts with adult cohorts in which sustained activity was associated with lower later healthcare costs [16–18]. Age is a likely part of the explanation. Follow-up ended in the late twenties, before many inactivity-related chronic conditions become common. At these ages, lower medication use can coexist with greater specialist contact, and a small number of high-cost events makes expenditure estimates imprecise. The public-total outcome is also deliberately narrow: it excludes person-level primary-care financing, fully private care, productivity losses, and nonmedical outcomes. It is an administrative payer aggregate, not total healthcare or societal cost.

The two sport measurements improve on a fixed baseline indicator, but they do not establish a continuous trajectory. The point estimate was lowest among students reporting sport at both assessments, yet it was not clearly different from the grade-8-only estimate. The grade-8-only category may contain students who stopped between waves, fluctuated in and out of sport, or were misclassified at one assessment; the grade-10-only category does not reveal when participation began. The appropriate interpretation concerns observed patterns at two school stages, not causal effects of starting, stopping, or maintaining sport.

Several limitations should be considered. Sport was measured only twice and without frequency, intensity, sport type, competitive level, or informal activity. The final cohort represented 43.5% of the source cohort and was somewhat younger, more female, and socioeconomically advantaged, although trimmed selection weighting produced similar estimates. Accordingly, the estimates are most directly generalizable to students who remained observable through both school assessments and had complete baseline information, rather than to the full 2008 grade-8 source cohort. The health modules lacked diagnosis, procedure, and drug-class detail, and fully private care was not observed. Most importantly, the analysis remains observational. Temporal ordering and extensive measured adjustment strengthen the prospective evidence but do not identify a causal effect.

These findings do not support a simple cost-saving account of adolescent sport. They instead point to differences in how young adults use the health system and highlight why the healthcare implications of adolescent sport should be evaluated across service sectors rather than inferred from prescriptions or aggregate expenditure alone. Future work should link repeated sport measures to diagnosis and drug-class

data and follow cohorts further into adulthood, when the balance between chronic disease, medication, injury, and specialist care may change.

Conclusions

Adolescent extracurricular sport participation was associated with distinct patterns of healthcare use through young adulthood. In an 11-year linked cohort of 23,283 Hungarian students, reporting sport at both assessments was associated with fewer dispensed prescriptions and primary-care contacts but more specialist outpatient visits than reporting sport at neither assessment. The inverse prescription association remained stable across sensitivity analyses and was also observed, with a smaller magnitude, in a separate cohort. No clear association with observable publicly reimbursed expenditure was detected. By combining repeated sport assessments with long, nearly complete administrative follow-up, this study documents how lower use in some healthcare sectors can coexist with higher use in others within the same cohort. Evaluations of adolescent sport should therefore examine the composition of subsequent healthcare use alongside expenditure, as neither a single service measure nor an aggregate spending total captures the full pattern of associations.

Abbreviations

Admin4 - Linked Administrative Panel Database

CI - confidence interval

HUF - Hungarian forint

NEAK - National Health Insurance Fund

OKM - National Assessment of Basic Competencies

PPML - Poisson pseudo-maximum likelihood

RECORD - REporting of studies Conducted using Observational Routinely-collected health Data;

STROBE - Strengthening the Reporting of Observational Studies in Epidemiology

Declarations

Ethics approval and consent to participate

This study was approved by the Research Ethics Committee of the Hungarian University of Sports Science (Magyar Testnevelési és Sporttudományi Egyetem Kutatásetikai Bizottsága), Budapest, Hungary (approval reference MTSE-KEB/No19/2026). The study was conducted in accordance with the principles of the Declaration of Helsinki and with applicable data-protection legislation, including Regulation (EU) 2016/679 (General Data Protection Regulation) and Hungarian Act CXII of 2011 on the Right of Informational Self-Determination and on Freedom of Information. Under the approved protocol, individual informed consent

was not sought, because the study is a secondary analysis of existing administrative records originally collected for public administrative purposes, the researchers had no access to direct identifiers, and contacting individuals would not have been feasible given the size of the cohort and the time elapsed since data collection. Data linkage was performed by authorized data controllers, and the analyses were conducted on pseudonymized data within the controlled environment of the ELTE Centre for Economic and Regional Studies Databank, under a data-access agreement between the Hungarian University of Sports Science and the Databank (permit EM-2025-0010) that sets out the conditions and rules governing the use of these data.

Consent for publication

Not applicable. The manuscript contains no identifiable individual-level information.

Availability of data and materials

The individual-level data are held by the ELTE Centre for Economic and Regional Studies Databank and contain sensitive administrative records. They are not publicly available and cannot be shared by the authors. Researchers may apply to the Databank for access, subject to project approval, data-protection requirements, and the relevant access rules. No separate study protocol or public analysis-code repository accompanies this submission; analyses were conducted in Stata 17 within the secure Databank environment.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received financial support from the Board of Trustees of the Hungarian University of Sports Science through the foundation responsible for maintaining the university, in the form of salary supplementation to the researchers involved. The funder had no role in the study design; data collection, management, analysis, or interpretation; preparation of the manuscript; or the decision to submit the manuscript for publication.

Authors' contributions

ÁK conceptualized the study, designed the record-linkage cohort, developed the exposure classification and the outcome definitions, curated and prepared the linked data, interpreted the findings, wrote the original draft, and was responsible for project administration and correspondence. TS supervised the project, contributed to the conceptualization and the framing of the research question and critically revised the

manuscript for important intellectual content. EK-G contributed to the interpretation of the findings in the context of adolescent sport participation, to the review of the literature, and to the review and editing of the manuscript. ZB and GC were responsible for the methodology and the empirical strategy, specified and estimated the Poisson pseudo-maximum likelihood models with school and calendar-year fixed effects, implemented the clustering and multiplicity procedures, carried out the secondary and sensitivity analyses, verified the reproducibility of the estimates, and drafted the statistical sections of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1.6 million participants. Lancet Child Adolesc Health. 2020;4:23-35. doi:10.1016/S2352-4642(19)30323-2.
2. World Health Organization. WHO guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020.
3. Bengtsson D, Svensson J, Wiman V, Stenling A, Lundkvist E, Ivarsson A. Health-related outcomes of youth sport participation: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2025;22:89. doi:10.1186/s12966-025-01792-x.
4. Crane J, Temple V. A systematic review of dropout from organized sport among children and youth. Eur Phys Educ Rev. 2015;21:114-131. doi:10.1177/1356336X14555294.
5. Tammelin T, Nayha S, Hills AP, Jarvelin MR. Adolescent participation in sports and adult physical activity. Am J Prev Med. 2003;24:22-28. doi:10.1016/S0749-3797(02)00575-5.
6. Belanger M, Giroux MA, Registe PPW, Gallant F, Jemaa S, Faivre P, et al. Adolescent physical activity profiles as determinants of emerging adults' physical activity. Int J Behav Nutr Phys Act. 2025;22:35. doi:10.1186/s12966-025-01732-9.
7. Kongsvold A, Skarpsno ES, Flaaten M, Logacjov A, Bach K, Nilsen TIL, et al. Adolescent sport and exercise participation and adult body composition and device-measured physical activity: the HUNT Study. Int J Behav Nutr Phys Act. 2025;22:29. doi:10.1186/s12966-025-01726-7.
8. Telama R, Yang X, Viikari J, Valimaki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med. 2005;28:267-273. doi:10.1016/j.amepre.2004.12.003.
9. Manz K, Krug S, Schienkiewitz A, Finger JD. Determinants of organised sports participation patterns during the transition from childhood to adolescence in Germany: results of a nationwide cohort study. BMC Public Health. 2016;16:939. doi:10.1186/s12889-016-3615-7.
10. Owen KB, Nau T, Reece LJ, Bellew W, Rose C, Bauman A, et al. Fair play? Participation equity in organised sport and physical activity among children and adolescents in high income countries: a

systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2022;19:27. doi:10.1186/s12966-022-01263-7.

11. Rittsteiger L, Hinz T, Oriwol D, Wäsche H, Santos-Hövener C, Woll A. Sports participation of children and adolescents in Germany: disentangling the influence of parental socioeconomic status. BMC Public Health. 2021;21:1446. doi:10.1186/s12889-021-11284-9.
12. Kirk SFL, Kuhle S, Ohinmaa A, Veugelers PJ. Health behaviours and health-care utilization in Canadian schoolchildren. Public Health Nutr. 2013;16:358-364. doi:10.1017/S1368980012002728.
13. Griffiths LJ, Cortina-Borja M, Tingay K, et al. Are active children and young people at increased risk of injuries resulting in hospital admission or accident and emergency department attendance? Analysis of linked cohort and electronic hospital records in Wales and Scotland. PLoS One. 2019;14:e0213435. doi:10.1371/journal.pone.0213435.
14. Kleppang AL, Hartz I, Thurston M, Hagquist C. Leisure-time physical activity among adolescents and subsequent use of antidepressant and hypnotic drugs: a prospective register linkage study. Eur Child Adolesc Psychiatry. 2019;28:177-188. doi:10.1007/s00787-018-1160-x.
15. Fahim M, Williams JVA, Ronksley PE, Fidler-Benaoudia MM, et al. Longitudinal study examining the relationship between physical activity and psychiatric hospitalizations in Canadian adolescents and young adults utilizing record linkage. Sci Rep. 2024;14:30281. doi:10.1038/s41598-024-81273-6.
16. Gomes GAO, Brown WJ, Codogno JS, Mielke GI. Twelve year trajectories of physical activity and health costs in mid-age Australian women. Int J Behav Nutr Phys Act. 2020;17:101. doi:10.1186/s12966-020-01006-6.
17. Coughlan D, Saint-Maurice PF, Carlson SA, Fulton J, Matthews CE. Leisure time physical activity throughout adulthood is associated with lower Medicare costs: evidence from the linked NIH-AARP Diet and Health Study cohort. BMJ Open Sport Exerc Med. 2021;7:e001038. doi:10.1136/bmjsem-2021-001038.
18. Codogno JS, Brown WJ, Turi-Lynch BC, de Morais LC, Lemes IR, Monteiro HL, et al. Primary health care costs associated with trajectories of physical activity over 10 years. J Phys Act Health. 2025;22:1013-1020. doi:10.1123/jpah.2025-0047.
19. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. Annu Rev Public Health. 2011;32:91-108. doi:10.1146/annurev-publhealth-031210-100700.
20. Thygesen LC, Ersbøll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. Eur J Epidemiol. 2014;29:551-558. doi:10.1007/s10654-013-9873-0.
21. Connelly R, Playford CJ, Gayle V, Dibben C. The role of administrative data in the big data revolution in social science research. Soc Sci Res. 2016;59:1-12. doi:10.1016/j.ssresearch.2016.04.015.
22. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H. Challenges in administrative data linkage for research. Big Data Soc. 2017;4:2053951717745678. doi:10.1177/2053951717745678.
23. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12:e1001885. doi:10.1371/journal.pmed.1001885.

24. Shaw RJ, Harron KL, Pescarini JM, Pinto Junior EP, Allik M, Siroky AN, et al. Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses. Eur J Epidemiol. 2022;37:1215-1224. doi:10.1007/s10654-022-00934-w.
25. Boza I, Pető R, Tir M. Hungarian administrative employer-employee data: the Admin4 database. KRTK-KTI Working Papers 2026/1. Budapest: ELTE Centre for Economic and Regional Studies; 2026.
26. KRTK Databank. Admin3 (H1) basic description and variable catalogue. Budapest: ELTE Centre for Economic and Regional Studies; 2023.
27. Hungarian Educational Authority. National Assessment of Basic Competencies: technical documentation and student background questionnaires, 2008 and 2010. Budapest: Hungarian Educational Authority.
28. Hungarian Central Statistical Office. Consumer price indices, annual data. Budapest: Hungarian Central Statistical Office. Available from: https://www.ksh.hu/stadat_files/ara/en/ara0001.html. Accessed 23 July 2026.
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology statement: guidelines for reporting observational studies. PLoS Med. 2007;4:e296. doi:10.1371/journal.pmed.0040296.
30. Santos Silva JMC, Tenreyro S. The log of gravity. Rev Econ Stat. 2006;88:641-658. doi:10.1162/rest.88.4.641.
31. Correia S, Guimarães P, Zylkin T. Fast Poisson estimation with high-dimensional fixed effects. Stata J. 2020;20:95-115. doi:10.1177/1536867X20909691.
32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289-300.

Figures

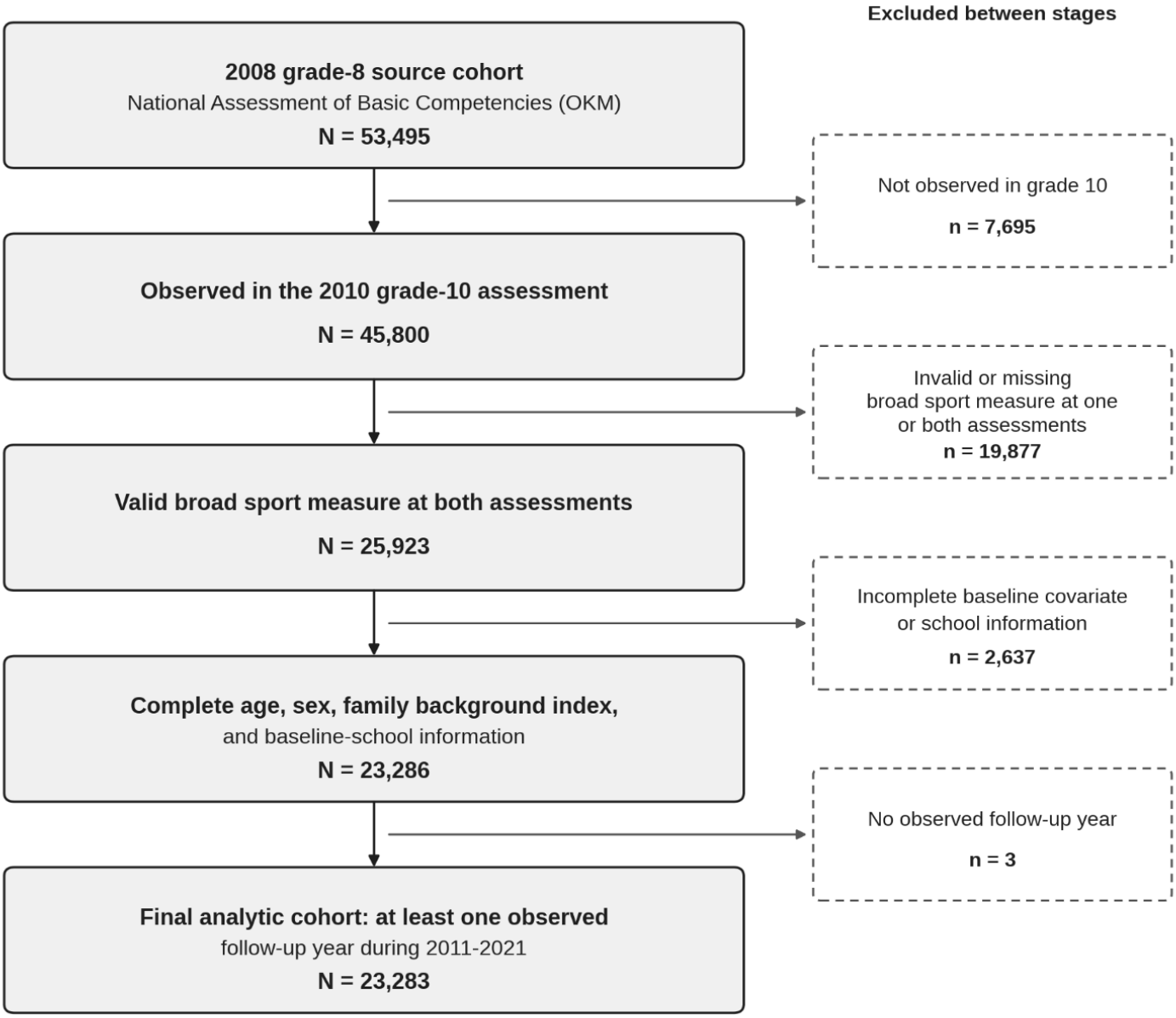


Figure 1

Participant flow from the 2008 grade-8 source cohort to the 2011-2021 analytic sample.

Note. Sequential exclusions reflect re-observation in the 2010 grade-10 assessment, availability of valid broad sport measures at both assessments, complete age, sex, family-background, and baseline-school data, and contribution of at least one observed follow-up year.

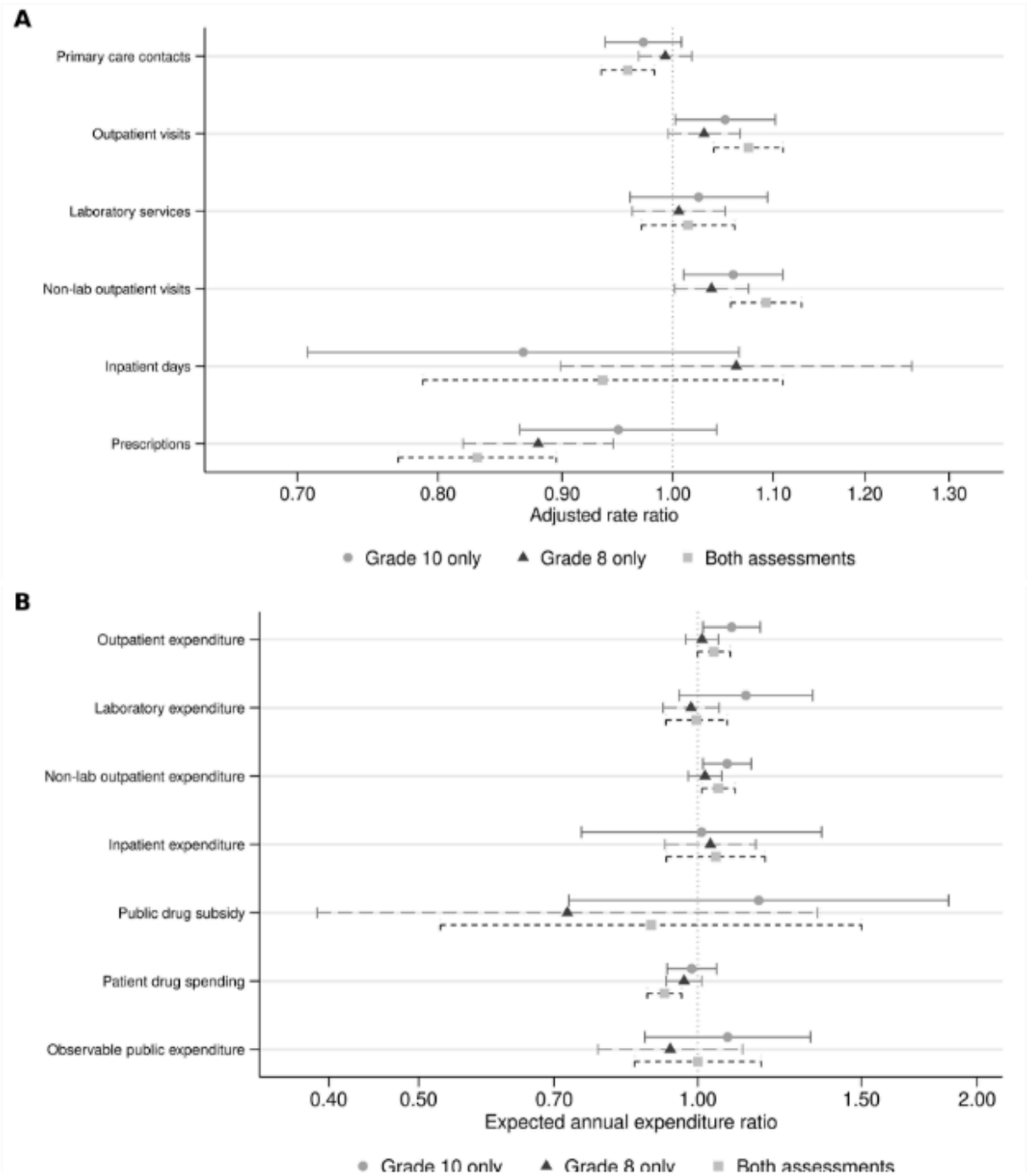


Figure 2

Adjusted associations between adolescent sport patterns and annual healthcare outcomes.

Note. Panel A shows adjusted rate ratios for utilization outcomes; panel B shows expected annual expenditure ratios. Points are estimates and bars are 95% confidence intervals. The reference group reported no regular extracurricular sport at either assessment; compulsory physical education was excluded. Models adjust for age, sex, family background, baseline-school fixed effects, and calendar-year fixed effects. Standard errors are clustered by person and baseline school. Expenditures are in constant 2021 HUF. Observable publicly reimbursed expenditure equals outpatient plus inpatient expenditure plus

public pharmaceutical subsidy and excludes person-level primary-care financing and fully private care. Horizontal axes are logarithmic.

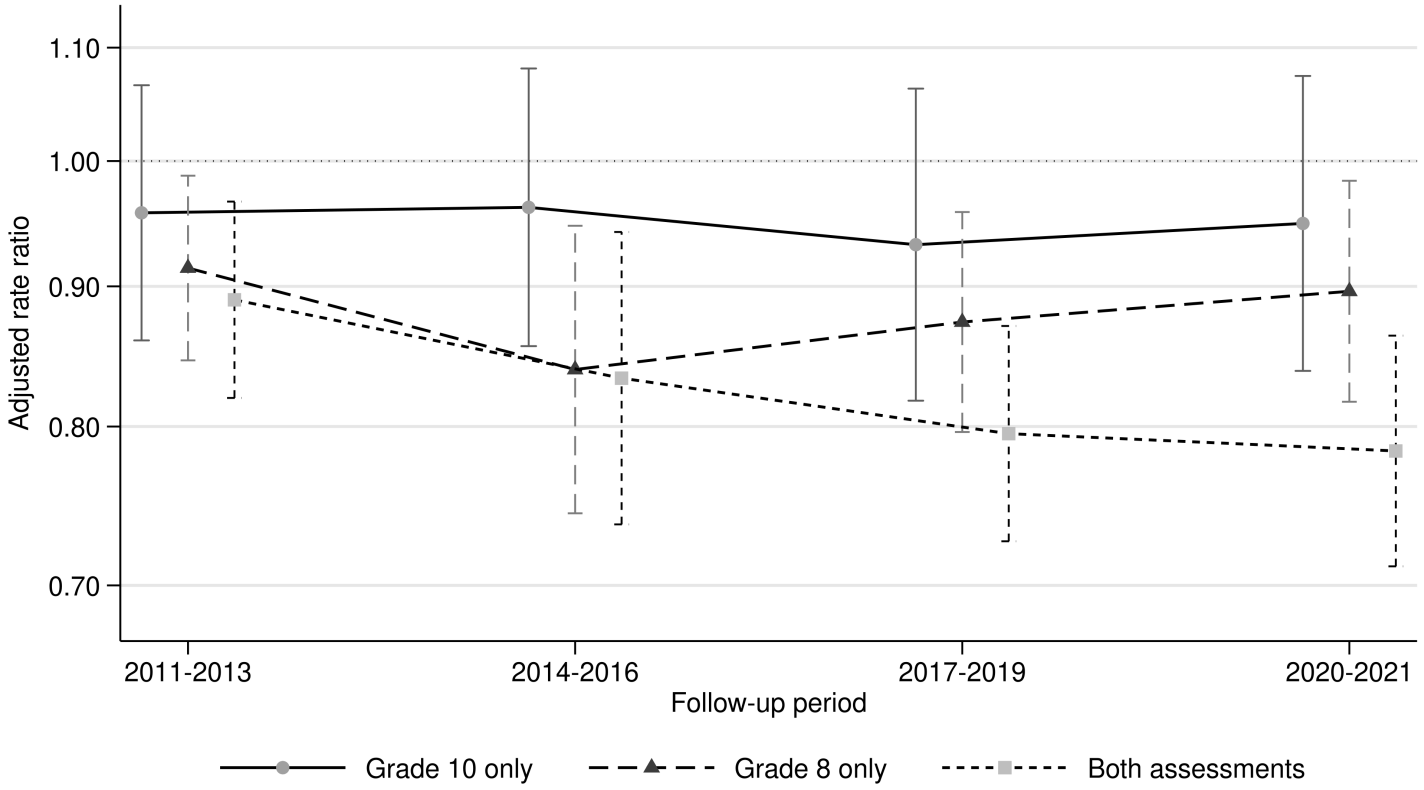


Figure 3

Adjusted prescription-rate ratios by follow-up period.

Note. Points are adjusted rate ratios and bars are 95% confidence intervals. The reference group reported no regular extracurricular sport at either assessment; compulsory physical education was excluded. Models adjust for age, sex, family background, baseline-school fixed effects, and calendar-year fixed effects. Standard errors are clustered by person and baseline school. The joint sport-pattern-by-period interaction test yielded p = 0.0473. The vertical axis is logarithmic.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2STROBEchecklist.docx
- Additionalfile3RECORDchecklistandlinkagesupplement.docx