Discussion Paper

Unit-level time-series multilevel models for small area estimation of sickness absence indicators

Harm Jan Boonstra and Jan van den Brakel

December 14, 2023

Contents

Abstract

The Netherlands Working Conditions Survey is an annual survey that measures, among other variables, various aspects of sickness leave of employees. Within the context of the Arbeidsmarkt Zorg en Welzijn (AZW) program, estimates are computed for several sickness leave indicators for a detailed breakdown of the AZW population of employees in the Human Health and Social Work Activities branch. In particular, estimates are computed for 28 regions, 10 AZW branches (as well as the non-AZW branch), 4 age classes, 3 employer size classes, as well as several cross-classifications of these individual breakdown variables. The desired high level of detail implies that direct estimates based on the survey weights are too inaccurate for many subpopulations of interest due to small sample sizes. Therefore, a small area estimation approach is developed, using data from the period 2014-2022, based on unit-level time-series multilevel models that account for all classifications of interest as well as for selectivity of the survey response with respect to the target population. The estimates based on the developed time-series models are compared with the direct estimates as well as with estimates based on cross-sectional multilevel models fitted to each year of survey data separately.

1 Introduction

The Netherlands Working Conditions Survey (Nationale Enquête Arbeidsomstandigheden or NEA, in Dutch) is an annual survey that measures (changes in) the working conditions of employees in the Netherlands. The survey is conducted by CBS (Statistics Netherlands) and TNO (Netherlands Organisation for Applied Scientific Research). The surveyed population consists of all employees from age 15 to (and including) 74 years who work in the Netherlands and who are registered as resident of the Netherlands, excluding those living in institutional households. For detailed information about the design and methodology of the NEA survey see van Dam et al. (2022). Starting with the reporting year 2022, the way data for the NEA are collected and processed differs in some respects from previous reporting years. As a result, the figures from 2022 may not be comparable in all cases to the figures through 2021. More information is available in the survey description of NEA.[1)] In particular, changes have been made to the questionnaire, sampling design, and weighting of NEA, see CBS and TNO (2023).

One of the topics covered by NEA is sickness absence. Within the Arbeidsmarkt Zorg en Welzijn (AZW) program there is a demand for detailed figures about sickness absence regarding the AZW subpopulation of employees in the Human Health and Social Work Activities branch. Desired figures are, among others, estimates of several sickness absence measures by region and AZW branch. The regional breakdown considered is defined by the so-called RegioPlus regions, a subdivision of the Netherlands in 28 regions. The AZW branch classification is a subdivision of the AZW sector into 10 branches. For the much bigger non-AZW sector, no sectorial subdivisions are considered in this project. The cross-classification of RegioPlus and AZW branch therefore defines a

1) See `https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksomschrijvingen/nationale-enquete-arbeidsomstandigheden--nea--`.

total of 308 = 28 × 11 subpopulations or domains. Other domains of interest are defined in terms of less detailed regional breakdowns such as province, AZW branch or the AZW/non-AZW split, as well as age class of the employee and size class of the employer.

Regular estimates based on NEA data are computed using the survey weights (van Dam et al., 2022). This can be done also for subpopulation estimates, where the weighted sums are restricted over the NEA data subset corresponding to each subpopulation. Such direct estimates, however, suffer from high variances in the case that the numbers of observations in (part of) the subpopulations become small. In such a case a model-based estimation methodology, generally known in official statistics as small area estimation (Rao and Molina, 2015), can help to obtain more accurate estimates. This usually entails using a multilevel model over all domains such that domain estimates also benefit to some extent from similar data in other domains. Here we consider the use of time-series multilevel models, which are fit on data from multiple instances of the NEA survey, in our case on data from 2014-2022. This way estimates can improve even more by exploiting similarities over time.

As part of the same AZW program a previous feasibility study of small area estimation of position in the job and current education was carried out using data from the Labour Force Survey (de Vries and Michiels, 2019). The same classification variables RegioPlus and AZW branch have been used in that study. The NEA survey has a much smaller number of annual observations than the Labour Force Survey, so that the need for a small area estimation method is even more urgent for NEA-based estimates at this level of detail. In a pilot study, small area estimates for a subset of sickness absence indicators have been computed based on unit-level multilevel models fitted to NEA 2019 data only. The evaluation in Boonstra et al. (2021) showed that for the most detailed figures the accuracy was not completely satisfactory, which was an important motivation to consider time-series multilevel models that use multiple years of NEA data. In the same pilot study it was also shown that using basic unit-level models with a single batch of random effects at the most detailed classification level (in this case RegioPlus × AZW branch) was suboptimal. In particular, models including in addition separate batches of random effects at the AZW branch and RegioPlus levels turned out to yield more plausible and less synthetic estimates, especially at the marginal AZW branch level.

The remainder of this paper is structured as follows. In Section 2, the survey design of the NEA and the available data sources considered for the small area estimation models are described. The multilevel models considered in this paper to produce small area estimates are described in Section 3. The resulting estimates based on the developed time-series multilevel models are presented in Section 4, including comparisons with direct estimates and cross-sectional model estimates. The paper concludes with a discussion in Section 5.

2 Data sources

2.1 The Netherlands Working Conditions Survey

The Netherlands Working Conditions Survey (Nationale Enquête Arbeidsomstandigheden or NEA, in Dutch) is an annual survey that measures (changes in) the working conditions of employees in the Netherlands. The annual sampling frames for the NEA survey consist of all registered employees between 15 and 75 years old. The size of the target population has gradually increased from 7,0 million to 8,0 million employees over the period 2014-2022. The sampling frame in each year contains many demographic variables, such as gender, age, migration background, region of residence (including province and RegioPlus), degree of urbanisation corresponding to the municipality of residence, etc. Also included are the design variables, i.e. the variables used in the sampling design of NEA, see van Dam et al. (2022). The design entails stratified sampling of employees, where the main stratification variable is a subdivision into 42 classes based on industry code. Besides, young employees (below 25) and persons with a non-western migration background are oversampled in NEA 2014-2021 to compensate for higher non-response rates among those groups. This oversampling has been discontinued in 2022. Each year, the NEA survey respondents are weighted so as to match the population distribution of several characteristics available in the sampling frame and other linked registrations. The weights are used to compute regular (direct) estimates, thus accounting for the stratification and oversampling used in the sampling design as well as for different nonresponse rates with regard to these characteristics.

The number of employees in the annual NEA response datasets varies between 38 thousand and 63 thousand over the 2014-2022 period. The target variables of interest relating to sickness absence[2)] are

1. Percentage absence time: the total number of absent days over the last twelve months divided by the total number of workable days.
2. Binary absence: whether an employee has been absent in the last twelve months.
3. Absence frequency: the number of absent periods in the last twelve months.
4. Number of absence days: the number of absent days in the last twelve months.
5. Duration of the last absence (not necessarily in the last twelve months): (1-4 days, 5-19 days, 20-209 days, 210 or more days).
6. Work-relatedness of the last absence (not necessarily in the last twelve months): (mainly work-related, partly work-related, not work-related, unknown).
7. The type of complaints experienced over the last absence (not necessarily in the last twelve months): (physical, psycho-social, other).
8. Most important reason that led to work-related complaints: (physical workload, psycho-social workload, other factors).

The data for the first variable are percentages, for the second variable binary indicators, and for the third and fourth variables counts. Variables 5, 6, 7 and 8 are categorical variables with more than two exclusive categories.

For the first 6 absence variables, official figures by RegioPlus and by AZW branch, but *not* by their cross-classification, have been published before on the Statistics Netherlands outputbase StatLine. However, since annual direct estimates are not sufficiently precise, these figures were averaged over the last 3 available years of NEA data. In this project we develop annual estimates for absence indicators corresponding to all 8 variables listed above, for an extensive set of subpopulations.

To be precise, the set of subpopulations for which annual estimates of the absence indicators are sought are listed in Table 2.1.

In the 'label' column of Table 2.1 'Total' refers to the full target population of all employees in the Netherlands, 'AZW broad' refers to the population of all employees in the AZW sector, 'AZW narrow' is defined as AZW broad minus the childcare branch, and

2) In the following whenever we speak of absence we mean absence due to sickness

	label	number of domains	group
1	Total, AZW broad, AZW narrow	3	total
2	AZW branch	11	AZW branch
3	Size class	3	total
4	Age class × (Total, AZW broad, AZW narrow)	5 × 3 = 15	total
5	Age class × AZW branch	5 × 11 = 55	AZW branch
6	Age class × Size class	5 × 3 = 15	total
7	Region × (Total, AZW broad, AZW narrow)	4 × 3 = 12	region
8	Province × (Total, AZW broad, AZW narrow)	12 × 3 = 36	region
9	RegioPlus × (Total, AZW broad, AZW narrow)	28 × 3 = 84	RegioPlus
10	Region × AZW branch	4 × 11 = 44	region-branch
11	Province × AZW branch	12 × 11 = 132	region-branch
12	RegioPlus × AZW branch	28 × 11 = 308	RegioPlus-branch

Table 2.1 Tables of domains of interest. The last column contains a grouping of the tables used in presenting some of the results in Section 4.

AZW branch is the subdivision of the AZW broad into 10 branches completed by the non-AZW sector, see Appendix B. Further, 'Size class' is a subdivision of the total population of employees into those that work at small, medium or large enterprises, and 'Age class' refers to the subpopulations of employees in the (partially overlapping) age groups 15-34, 35-54, 55-74, 35-44 and 45-54 years old. Three hierarchically related regional subdivisions are of interest: 'Region' ('landsdeel' in Dutch) is a subdivision of the Netherlands into 4 parts, 'Province' is the next more fine-grained subdivision into 12 regions, and 'RegioPlus' is the most detailed subdivision considered, further subdividing the provinces into 28 regions in total. Overall, estimates are computed for each year in the 2014-2022 period, for each of the 8 indicators (of which the last four have multiple categories), and for each of the in total 718 subpopulations considered.

Due to item non-response, the number of available observations in each year differs between the eight target variables. Binary absence has the fewest missings, about 50 per year on average. Duration of the last absence has the largest amount of item-non-response with an average of more than 2500 missings per year. Note that for target variables 5 - 8 fewer data are available anyway, because not all employees are eligible: these variables are ineligible for employees who have never had a period of sickness leave, where variable 8 is in addition ineligible for those who have had only non-work-related absences.

2.2 Direct estimates

Direct estimates based on NEA are computed using the NEA weights, which have been derived using a multiplicative weighting method that matches weighted NEA means to unweighted population means for background characteristics gender, age, migration background, industry branch, province, degree of urbanisation and educational attainment (van Dam et al., 2022). For NEA 2022 the weighting scheme, along with the sampling design, has changed (CBS and TNO, 2023), and now uses income in classes instead of education, and in addition household type and type of contract (temporary/permanent).

For each year we compute the direct estimates and corresponding variance estimates for all target variables and domains of interest, so that they can be compared to the model-based small area estimates discussed later. The direct estimate for the population mean in a certain domain d of interest is computed as

$$\hat{\bar{Y}}_d = \frac{\sum_{i \in s_d} w_i y_i}{\sum_{i \in s_d} w_i}, \tag{1}$$

where y is one of the absence variables of interest, s_d is the set of employees in domain d for which y is observed, and w_i is the NEA weight for employee i. Corresponding variance estimates are computed as

$$v(\hat{\bar{Y}}_d) = \frac{1}{n_d(n_d - 1)} \sum_{i \in s_d} (y_i - \bar{y}_d)^2, \tag{2}$$

where $\bar{y}_d$ is the mean of y within s_d and n_d is the number of employees in s_d. Note that weights are not used in these variance estimates. A slightly refined variance estimate would include a variance inflation factor due to the variation in weights within each domain, but we have checked that this effect is quite small.

Regarding numbers of observations, AZW branch response sizes range from an annual average of less than 200 (GPs and health centers, and Youth care) to more than 2000 (Nursing, care and homecare), and over 40000 in the non-AZW remainder. The RegioPlus response sizes are more balanced, ranging from 700 to 3400 per year on average over the 2014-2022 period. For the cross-classification of AZW branch and RegioPlus the response sizes range from 0 to 3000 per year on average. In particular, over the period 2014-2022 it happens 38 times that a combination of AZW branche and RegioPlus has no observations (and therefore undefined direct estimates and standard errors), and 52 times that there is only a single observation (hence undefined standard error estimates according to (2)). Note that the actual number of observations available for each indicator can be smaller due to item-non-response and ineligibility. Given these numbers, it can be expected that model-based small area estimates on average will be much more accurate than the direct estimates.

2.3 Additional data sources for small area estimation

The compiled sampling frames already contain many covariates that can be used for weighting/modelling or analyses. Sampling frame covariates that we use for modelling are gender, age class, migration background, stratum including oversampled educational branches, as well as the regional variables. Of these, stratum, age and migration status are used in the NEA sampling design. See Appendix A for an overview of all covariates used.

To enrich the sampling frame with further relevant covariates, data from two other registrations have been matched to the sampling frame:

- The Municipal Base Administration (Basisregistratie Personen or BRP in Dutch). From this register, household type was added to the sampling frame. We use the BRP versions corresponding to the second quarter of each year, since they match best with the sampling frames, which are compiled around the same time.
- The jobs register of the second quarter of each year. This register contains information on jobs of employees. For employees with multiple jobs we choose the one with most working hours, which is also the job that most NEA questions refer to. From this source several variables are used: income, size class of the employing enterprise, type of job contract, and industry code.

All covariates used are categorical, see Appendix A. For all registers, and in particular those with reference periods furthest away from the sampling frame's reference dates, it is the case that a certain percentage of sampling frame employees don't match. For these non-matching units we use the category 'unknown', which subsequently is collapsed with one of the other categories, usually the largest one, to avoid very small classes.

The AZW branch classification variable is derived from the industry code and collective labour agreement variables of the jobs register. In this case non-matching employees are assigned to the remainder non-AZW class, which contains more than 80% of the employees.

3 Unit-level models for small area estimation

In a previous study, (Boonstra et al., 2021) applied unit-level multilevel models to estimate a subset of absence indicators based on the NEA survey data of 2019. We refer to such models applied to a single year of survey data as cross-sectional models, in contrast to time-series models which simultaneously model data over a range of years. It was noted that for the most detailed figures considered, at the cross-classification level of RegioPlus and AZW branche, the accuracy of the cross-sectional model estimates was not yet satisfactory. Therefore, in order to improve the estimation accuracy we here consider modelling multiple years of NEA data using a time-series extension of unit-level multilevel models.

The best known unit-level model in small area estimation is the Battese-Harter-Fuller model, also known as nested error regression model, or simply as basic unit-level model (Battese et al., 1988; Rao and Molina, 2015). It is a linear Gaussian multilevel model with a single batch of random intercepts, appropriate for modelling data with a single detailed classification dimension of interest. For the absence indicators, there is interest in estimates at several detailed levels, notably by RegioPlus, AZW branche as well as their cross-classification. For that purpose (Boonstra et al., 2021) already considered more general unit-level multilevel models with multiple batches of random effects corresponding to the classification levels of interest. Besides, since the different absence indicators are based on different kinds of data, non-Gaussian multilevel models, such as from the binomial and negative-binomial families, were used as well.

Modelling multiple years of survey data, in our case NEA data of 2014-2022, asks for a time-series extension of the cross-sectional multilevel models. As time, i.e. year, becomes another dimension of interest, it should be sufficiently accounted for in the model, including in several random effect interaction terms, such that the model can partially pool data over both time and other dimensions of interest. In addition, for random effects involving the time dimension, we usually choose covariance specifications that account for their natural ordering.

In the remainder of this section we first discuss the general setup of the multilevel models including the random effect specification, and after that lay out the specific fixed and random effect terms selected for small area estimation of the various absence indicators.

3.1 General unit-level time-series multilevel models

Let y denote one of the target variable vectors, and y_i the observed value for employee i. We denote the length of y by n, which is the number of rows of the combined NEA 2014-2022 dataset minus the (relatively) small number of missing values due to item-nonresponse. Let X be the $n \times p$ design matrix for a set of p covariates selected for inclusion in the model. The multilevel models considered take the generalized linear additive form

$$\begin{aligned} y_i &\overset{\text{ind}}{\sim} f(\mu_i, \phi) \\ g(\mu_i) = \eta_i &\equiv X_i\beta + \sum_{\alpha} Z_i^{(\alpha)} v^{(\alpha)} , \end{aligned} \tag{3}$$

where f is a probability distribution depending on a data-dependent mean μ_i and an optional scale or dispersion parameter ϕ, that we allow to vary by year in case of a Gaussian variance parameter. Further, g is a link function that links the mean μ_i to the linear predictor η_i. The latter is defined in terms of the covariate matrix X, with X_i denoting its ith row, and associated regression or fixed effects β, as well as a set of random effect design matrices $Z^{(\alpha)}$ of dimension $n \times q^{(\alpha)}$ and corresponding random effect vectors $v^{(\alpha)}$ of size $q^{(\alpha)}$. Here α runs over the different random effect terms used in the model. In the models considered we use several random effect terms, associated with the classification levels of interest including cross-sectional RegioPlus × AZW branch intercepts and several interactions with year and other classification variables of interest such as age class. The fixed effects part of the model includes effects associated with important covariates, which are explanatory for either the missing-data mechanism (consisting of both the sampling design and an unknown response mechanism) or the target variables or both.

Several data distributions are considered, suitable for the different target variables. For many of the variables a linear Gaussian model has been attempted where f denotes a normal distribution with mean μ_i and variance $\phi = \sigma^2$. In this case the link function used is always the trivial identity function. In the end, the linear Gaussian model is only selected as the best among the tried model families for target variables 1 (percentage absence time) and 3 (absence frequency), see the list of target variables on page 5. For target variable 2, binary absence, a binomial/Bernoulli model is used with logistic link, i.e. $g(\mu_i) = \log \frac{\mu_i}{1-\mu_i}$. For variable 4, number of absence days, we use a negative binomial distribution, with a logarithmic link function. In that case a dispersion parameter ϕ is allowed to be inferred from the data. For the remaining variables 5,6,7 and 8 we use multinomial logistic time-series multilevel models as these variables are defined in terms of multiple (> 2) categories. For the multinomial models (3) should be interpreted as being defined per category with the sum-to-one constraint implicitly imposed, see the discussion below in Subsection 3.5.

A Bayesian approach of model fitting and prediction is taken. In particular we use Markov Chain Monte Carlo simulation to fit the models, as discussed further in Subsection 3.6. The vector β of fixed effects is assigned a prior $\beta \sim N(0, 100 I_p)$, which is only very weakly informative given the scales of the target variables and covariates (all categorical), and therefore easily overwhelmed by the information in the data. In the case of a linear Gaussian model we allow the residual variance parameter to vary by year, assigning them independent inverse chi-squared priors with degrees of freedom parameter equal to 1. In the case of a negative binomial model a single overall dispersion parameter is used and assigned a chi-squared distribution with 1 degree of freedom.

The random effect vectors $v^{(\alpha)}$ for different α are assigned independent prior distributions. To describe the general prior for each vector $v^{(\alpha)}$ of random effects, we suppress superscript α from now on. Each random effect vector v is assumed to be distributed as

$$v \sim N(0, A \otimes V)\,, \tag{4}$$

where V and A are $d \times d$ and $l \times l$ covariance matrices, respectively, and $A \otimes V$ denotes the Kronecker product of A with V. The total length of v is ld, and these coefficients may be thought of as corresponding to d effects allowed to vary over l levels of a factor variable, e.g. intercepts ($d = 1$) varying over the $l = 28 \times 11 = 308$ levels of the interaction of RegioPlus and AZW branch. The covariance matrix A describes the covariance structure among the levels of the factor variable, and is assumed to be known. Instead of covariance matrices, precision matrices $Q_A = A^{-1}$ are actually used, because of computational efficiency (Rue and Held, 2005). For multiple varying effects where $d > 1$, the covariance matrix V among the d varying effects is parameterized in one of three different ways:

– an unstructured, i.e., fully parameterized covariance matrix
– a diagonal matrix with unequal diagonal elements
– a diagonal matrix with equal diagonal elements.

The following priors are used for the parameters in the covariance matrix V:

– In the case of an unstructured covariance matrix the scaled-inverse Wishart prior is used as proposed in O'Malley and Zaslavsky (2008) and recommended by Gelman and Hill (2007).
– In the case of a diagonal matrix with equal or unequal diagonal elements, half-Cauchy priors are used for the standard deviations. These priors are also used for scalar standard deviation parameters in case $d = 1$. Gelman (2006) demonstrates that these priors are better default priors than the more common inverse gamma priors for random effects' variance parameters.

3.2 Linear model

None of the absence variables conforms to a normal distribution. All target variables considered are discrete, even the absence percentage since it is defined as the number of absence days divided by the number of workable days in the last twelve months. Besides, all non-categorical variables are bounded below by 0 and above by either 1 (or 100% on a percentage scale) or the number of workable days. The absence frequency actually has an apparent cut-off at 40 absence periods.

Nevertheless, a linear Gaussian multilevel model is a convenient base model to fit, and sometimes works surprisingly well for the purpose of predicting population totals or means, such as is the case for small area estimation. For example, Boonstra et al. (2007) conducted a simulation study based on unemployment data from the Dutch Labour Force Survey, from which they conclude that for the task of estimating municipal unemployment fractions a linear multilevel model performs similarly to a logistic multilevel model tailored to the binary unemployment data.

For the linear models considered, equation (3) becomes

$$y_{ti} = X_{ti}\beta + \sum_{\alpha} Z_{ti}^{(\alpha)} v^{(\alpha)} + \epsilon_{ti} \quad \text{with} \quad \epsilon_{ti} \overset{\text{ind}}{\sim} N(0, \sigma_t^2)\,, \tag{5}$$

where subscript t runs over the years 2014-2022 and subscript i refers to the NEA survey respondents, and runs from 1 to n_t, the total number of respondents in year t.

3.3 Binomial model

The second target variable, binary absence, is a categorical variable with just two classes. For such variables a binomial distribution specialized to binary data, known also as Bernoulli distribution, is the most obvious distribution to use. The link function, linking the distribution's mean, i.e. the probability of a 'success', to the linear predictor, is commonly taken to be the inverse of the logistic function. The resulting model can be viewed as a multilevel generalization of logistic regression.

In this case, equation (3) becomes,

$$\begin{aligned} y_{ti} &\overset{\text{ind}}{\sim} Be(p_{ti}) \\ \log \frac{p_{ti}}{1-p_{ti}} &= X_{ti}\beta + \sum_{\alpha} Z_{ti}^{(\alpha)} v^{(\alpha)} , \end{aligned} \tag{6}$$

where $Be(p_{ti})$ denotes the Bernoulli distribution with parameter p_{ti}, i.e. $y_{ti} = 1$ with probability p_{ti} and $y_{ti} = 0$ with probability $1 - p_{ti}$.

3.4 Negative binomial model

For the absence frequency and number of absence days, both linear and negative binomial multilevel models are applied. The latter are generally more suitable for count data. Another popular probability distribution for count data is the Poisson distribution, which has the property that its mean and variance parameters are equal. The negative binomial model is more general in that it allows for a variance that is larger than the mean, i.e. overdispersion. We allow a single global dispersion parameter r of the negative binomial distribution to be inferred from the data.

In this case, equation (3) becomes,

$$\begin{aligned} y_{ti} &\overset{\text{ind}}{\sim} NBin(r, p_{ti}) \\ \log \frac{p_{ti}}{1-p_{ti}} &= X_{ti}\beta + \sum_{\alpha} Z_{ti}^{(\alpha)} v^{(\alpha)} , \end{aligned} \tag{7}$$

where $NBin(r, p_{ti})$ denotes the negative binomial distribution with dispersion parameter $r > 0$ and probability parameter p_{ti}, defined by the probability mass function

$$p(y_{ti}|r, p_{ti}) = \binom{y_{ti} + r - 1}{y_{ti}} (1 - p_{ti})^r p_{ti}^{y_{ti}} . \tag{8}$$

For r a positive integer, the negative binomial distribution is the distribution of the number of successes until a certain number r of failures has occurred. Its mean is $\mu_{ti} = \frac{rp_{ti}}{1-p_{ti}}$, and its variance $V(y_{ti}) = \frac{rp_{ti}}{(1-p_{ti})^2} = \mu_{ti}(1 + \frac{\mu_{ti}}{r})$. Note that smaller r means more overdispersion compared to the Poisson distribution. As r goes to infinity in such a way that the mean approaches a constant value, the distribution approaches the Poisson distribution. For small values of r the negative binomial distribution can fit data with an excess number of zeros, which can be useful for the absence variables as a large fraction of respondents indicate that they have not been absent over the previous 12 months.

3.5 Multinomial model

Target variables 5 to 8 are categorical. This corresponds to a special case of multinomial data where there is only a single trial. So for a categorical target variable with K categories, one and only one of y_{tik} for $k = 1, \ldots, K$ equals 1, whereas the others are 0.

The logistic multinomial multilevel model is specified as

$$
\begin{aligned}
\mathbf{y}_{ti} &\overset{\text{ind}}{\sim} Multinom(n_{ti}, \mathbf{p}_{ti}) \\
p_{tik} &= \frac{\exp(\eta_{tik})}{\sum_{k'=1}^{K} \exp(\eta_{tik'})}, \\
\eta_{tik} &= X_{ti}\beta_k + \sum_{\alpha} Z_{ti}^{(\alpha)} v_k^{(\alpha)},
\end{aligned}
\tag{9}
$$

where in the first line the notation in bold of $\mathbf{y}_{ti}$ for the target variable and $\mathbf{p}_{ti}$ for the probability parameters indicates the multivariate nature of this model. Note that we assume a common set of design matrices X and Z for all categories, whereas all fixed and random coefficients are category-specific. For categorical data, as is the case here, we have $n_{ti} = 1$ for all t, i.

For computational reasons we follow Linderman et al. (2015) in applying a so-called stick-breaking transformation to model (9), in order to represent it as $K - 1$ independent binomial distributions:

$$
\begin{aligned}
&\tilde{n}_{tik} = n_{ti} - \sum_{k'<k} y_{tik'}, \qquad \tilde{p}_{tik} = \frac{p_{tik}}{1 - \sum_{k'<k} p_{tik'}}, \\
&\Rightarrow \quad p(\mathbf{y}_{ti}|\mathbf{p}_{ti}) = \prod_{k=1}^{K-1} Binom(y_{tik}|\tilde{n}_{tik}, \tilde{p}_{tik}).
\end{aligned}
\tag{10}
$$

In the case of categorical data ($n_{ti} = 1$) this further reduces to a product of Bernoulli distributions. So instead of p_{tik}, we model the $\tilde{p}_{tik}$ as

$$
\log \frac{\tilde{p}_{tik}}{1 - \tilde{p}_{tik}} = X_{ti}\tilde{\beta}_k + \sum_{\alpha} Z_{ti}^{(\alpha)} \tilde{v}_k^{(\alpha)}. \tag{11}
$$

For $\tilde{\beta}_k$ and $\tilde{v}_k$ the priors as discussed in Subsection 3.1 are used. In particular, it is possible to allow for correlations among the random effects for different categories by using a full covariance matrix among the $\tilde{v}_k$ with $k = 1 \ldots K - 1$.

3.6 Estimation of the multilevel models

The models are fitted using Markov Chain Monte Carlo (MCMC) sampling, in particular the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). The full conditional posterior distributions used by the Gibbs sampler are all known distributions that are easy to sample from. For the binomial, multinomial and negative binomial unit-level models we use the data augmentation approach of Polson et al. (2013), in which the binomial-logistic likelihood is represented as a scale-mixture of normal distributions. In the negative binomial case the data augmentation approach of Zhou et al. (2012); Zhou and Carin (2015) results in a closed-form full conditional posterior for the dispersion parameter.

The MCMC simulations are run in R (R Core Team, 2015) using package `mcmcsae` (Boonstra, 2021). This package allows block Gibbs sampling where all coefficients, both fixed and random, are updated together, resulting in an efficient sampling algorithm. In particular, for the multinomial multilevel models block sampling of the coefficients corresponding to different categories is possible due to the use of the stick-breaking representation (Linderman et al., 2015). The Gibbs samplers implemented also use multiplicative redundant parameterization (Gelman et al., 2008) for random effects, further improving the overall sampling efficiency.

The Gibbs sampler is run in parallel for three independent chains with randomly generated starting values. For the selected models we use a burn-in of 1000 iterations, and then 2000 iterations of which only the parameter draws for every fourth iteration are retained. This yields $3 * 2000/4 = 1500$ samples as a basis for computing the estimates and standard errors. The thinning of the 2000 draws to 500 in each chain is done to reduce memory usage and to speed up the unit-level prediction for the large non-observed part of the population in each year. The multinomial models for target variables with K classes are about K times slower to fit, so as a compromise we use half as many iterations to speed up the computations, i.e. a burn-in of 500, then 1000 iterations where for each fourth iteration the parameter draws are stored, leaving 750 draws for inference.

The convergence of the MCMC simulations is assessed using trace and autocorrelation plots as well as the Gelman-Rubin potential scale reduction factor (Gelman and Rubin, 1992), which diagnoses the mixing of the chains. For the simulations of the selected models almost all potential scale reduction factors are below 1.02 (most very close to 1.00), and only for a few model parameters the potential scale reduction factor is above 1.05. The latter does not happen for the models for binary absence and absence frequency, but happens a few times for the models for the other variables, especially for the multinomial models. The maximum encountered potential scale reduction factor for a model parameter is 1.17 for the model for the duration of the last absence in four categories. We have also checked that the potential scale reduction factors for all predictive quantities of interest, i.e. for all indicators for all domains, are below 1.05, with most of them very close to 1.00, and that the estimated Monte Carlo standard errors are small compared to the posterior standard deviations.

3.7 Choice of fixed and random effects

For the regular production of estimates based on the NEA survey, a weighting procedure is applied each year to account for different inclusion and response probabilities. For the purpose of this study we employ unit-level models in which there is no obvious place for survey weights. In order to account for the survey design and unequal response probabilities, design characteristics such as strata and other covariates used in the NEA weighting schemes are also included as covariates in the unit-level models. Besides design and response related covariates, we also try to include additional covariates in order to increase the explained part of the response variables' variation.

Many models of the form (3) have been fitted to the various target variables. For the comparison of models using the same input data and the same distribution we make use of the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), the Widely Applicable Information Criterion or Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010, 2013) as well as the approximate Leave-One-Out cross-validation Information Criterion (LOOIC) (Vehtari et al., 2017). These model criteria take into account both model fit and model complexity through the effective number of model parameters, and are relatively easy to compute based on the MCMC simulation results. They are more suitable for comparing multilevel models than the more traditional AIC criterion, as the latter does not account for the random effects' prior specification.

The model adequacy of the selected models has also been evaluated using posterior predictive checks. This implies that replicate datasets simulated from the posterior predictive distribution are compared with the originally observed data to study systematic discrepancies and to evaluate how well the selected model fits the observed data (Gelman et al., 1996).

An invaluable part of the model assesment is the graphical comparison of the model-based estimates to direct estimates based on the survey weights, at various aggregation levels. In particular, at high aggregation levels the model-based estimates are expected to be close to the corresponding direct estimates, particularly when averaged over time. To make sure that the aggregated model-based estimates exactly equal the regular annual direct estimates for the complete population of all employees in the Netherlands, a benchmarking step is added at the end of the workflow. This only causes minor shifts in the underlying model-based estimates, but ensures that at the overall level the estimates are consistent with regular figures that have already been published.

We have not attempted to optimize the set of selected covariates and random effects for each target variable separately. Instead a practical choice has been made to use the same set of covariates and random effects for each target variable. This choice seems reasonable because all estimates are based on the same survey data (apart from a small amount of item non-response), the same set of sub-populations is of interest for all indicators, and the target variables themselves all measure aspects of sickness absence and are quite strongly related. The models still differ regarding the assumed family of sampling distributions suitable for the type of data.

The model parameters in (3) are separated in fixed and random effects. After extensive examination of different models, the following fixed effects components are considered in the selected models for all response variables:

$$\begin{aligned}
&sex * (ageclass6 + background) + sizeclass3 * ageclass3 + \\
&AZW * (sex + ageclass6 + nonwestern + income + hhtype + \\
&\qquad sizeclass3 + Region + contract) + \\
&Stratum + Year * Oversampled + Urban + \\
&sizeclass9 + RegioPlus + AZWbranch
\end{aligned} \tag{12}$$

See Appendix A for an overview of the covariates used. It is understood that terms like $sex * ageclass6$ in (12) include both main and interaction effects. The variable AZW is a simple binary indicator variable for whether an employee is working in the AZW sector or not. Interactions of AZW with some important covariates have been included as fixed effects since the AZW sub-population is the population of main interest in this study. Note also that $Year$ is defined here as a categorical variable with a category for each of the years in 2014 to 2022.

Included in covariate model (12) are, at least approximately, all variables that are used in the NEA sampling design and the NEA weighting scheme. Before 2022 the NEA weighting scheme included educational level as a covariate, which in 2022 has been replaced by quantile-based income groups. We have found income to be the more relevant covariate, so we have included $income$, but not educational level in (12).

Note that the only interaction with $Year$ in (12) is the term $Year * Oversampled$. This term is included because the oversampling of educational branches has changed over time. For other variables the changes over time regarding oversampling or response propensities are deemed smaller, and therefore fixed effect interactions with year are omitted. However, important interactions with $Year$ are mostly added as random effects, as will be discussed next.

For the selection of random effect components the most important considerations concern the aggregation levels of interest. For estimation at a particular aggregation level it is desirable to include effects in the model for all underlying classes. For example, if one is interested in annual estimates by AZW branch, AZW branch effects for each year should ideally be included in the model. Otherwise the estimates tend to become synthetic in the sense that differences between AZW branches within a year or developments over time within a branch are underestimated unless other covariates can completely account for such differences, which is rare. However, in the small area estimation context where the number of observations in many of these classes can be quite small or even zero, these effects cannot be included as fixed effects as that would result in very noisy effect estimates. Therefore such effects are modelled as random effects, to obtain a better bias-variance tradeoff. For this application it means that we include random effects for the interaction of RegioPlus and AZW branch, as well as interactions of RegioPlus and AZW branch with Year to account for time-variation. In case of time-varying effects it makes sense to use a covariance structure that accounts for the natural ordering of the years. We do this by incorporating the precision matrix corresponding to a random walk, see e.g. Rue and Held (2005), in the definition of A^{-1} in (4). A further choice is whether only intercepts or also other covariate effects are allowed to vary over the classes of the aggregation levels. In the case of multiple varying effects there is a choice between scalar, diagonal or full covariance matrix V in (4).

The linear predictor specification in (3) used in all selected models can be expressed as

$$\eta_{ti} = x'_{ti}\beta + v_{ts}^{\text{stratum}} + z'_{ti}u_t^{\text{dyn}} + v_{rb} + q'_{ti}u_{tr}^{\text{AZW}} + v_{ba}^{\text{age}} + u_{tba}^{\text{agedyn}} + w_{trb}\,, \tag{13}$$

where the first term involves the fixed effects β and all the other terms describe the random effect part of the model. Here random effects are denoted by u, v, w whereas q and z are covariate vectors. Note that we use u for random effects that define autocorrelated trend components over time. The following list describes all independently specified random effect terms in turn:

1. The term v_{ts}^{stratum} denotes random intercepts by year and stratum, so $v_{ts}^{\text{stratum}} \overset{\text{iid}}{\sim} N(0, (\sigma^{\text{stratum}})^2)$ for all strata s and years t. Note that the subscript s in (13) denotes the stratum to which employee i in year t belongs. A more precise subscript notation in (13) would use $s[ti]$ to indicate this relationship, but we opt for the more compact notation.
2. The term $z'_{ti}u_t^{\text{dyn}}$ describes the contribution of a set of dynamic regression coefficients. Here z_{ti} is a vector of covariates whose effect vectors u_t^{dyn} are allowed to vary over time according to the (first-order) random walk specification $u_{j,t}^{\text{dyn}} - u_{j,t-1}^{\text{dyn}} \overset{\text{ind}}{\sim} N(0, (\sigma_j^{\text{dyn}})^2)$ where j runs along the covariates selected for this term. For identifiability reasons the effects are constrained to satisfy $\sum_t u_{j,t}^{\text{dyn}} = 0$ for each j where the sum is taken over the years 2014 - 2022. In the selected models, we use the covariates corresponding to the specification

$$\begin{aligned} &AZW * (sizeclass3 + sex + ageclass6 + income + hhtype + \\ &\quad nonwestern + Region + contract) \end{aligned} \tag{14}$$

Each (dummy) covariate of (14) has its own variance parameter $(\sigma_j^{\text{dyn}})^2$, except that the same variance parameter is used for the two classes of AZW.

3. The term v_{rb} denotes random intercepts by RegioPlus and AZW branch, so $v_{rb} \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ for all RegioPlus regions r and AZW branches b.
4. The term $q'_{ti} u_{tr}^{\text{AZW}}$ defines regional trends over time. Here q_{ti} denotes the 2-vector with binary indicators for belonging to the AZW sector or not, and u_{tr}^{AZW} are the corresponding (2-vector) effects for each year t and each RegioPlus region r. These effects are modelled as random walks over the years, i.e. $u_{t,r}^{\text{AZW}} - u_{t-1,r}^{\text{AZW}} \overset{\text{iid}}{\sim} N\left(0, \begin{pmatrix} (\sigma_1^{\text{AZW}})^2 & \rho^{\text{AZW}} \\ \rho^{\text{AZW}} & (\sigma_2^{\text{AZW}})^2 \end{pmatrix}\right)$. Note that there are only two variance parameters here, $(\sigma_j^{\text{AZW}})^2$ for $j = 1, 2$ corresponding to the AZW population and the non-AZW population. The following sum-to-zero constraints are imposed for identifiability: $\sum_t u_{t,r,j}^{\text{AZW}} = 0$ for all r, j.
5. v_{ba}^{age} denote random intercepts by AZW branch and $ageclass3$, so $v_{ba}^{\text{age}} \overset{\text{iid}}{\sim} N(0, (\sigma^{\text{age}})^2)$ for all AZW branches b and age classes a.
6. The term u_{tba}^{agedyn} defines time trends by AZW branch and $ageclass3$, where the dependence over time is modelled using random walks: $u_{t,b,a}^{\text{agedyn}} - u_{t-1,b,a}^{\text{agedyn}} \overset{\text{ind}}{\sim} N(0, (\sigma^{\text{agedyn}})^2)$ with sum-to-zero constraints $\sum_t u_{t,b,a}^{\text{agedyn}} = 0$ for all b, a.
7. w_{trb} denote random intercepts by year, RegioPlus and AZW branch, so $w_{trb} \overset{\text{iid}}{\sim} N(0, \sigma_w^2)$ for all years t, detailed regions r and AZW branches b.

For the multinomial models there is a linear predictor (13) for each category except the last. This means that the overall model contains $K - 1$ versions of each parameter in (13). In addition, we allow the random effects of different categories to be correlated.

3.8 Cross-sectional multilevel model

For comparison purposes we also compute small area estimates based on a cross-sectional unit-level model, i.e. a model that is estimated using data for each year separately. Such a model cannot borrow strength over time, but can still borrow strength over the other dimensions, notably AZW branche and RegioPlus.

The cross-sectional models that we use for the 8 absence indicators are based on the same sampling distributions as used for the time-series models. Moreover, the fixed and random effects are selected to resemble as far as possible those selected in the time-series model, with some fine-tuning especially regarding the fixed effects part to prevent too much overfitting. This way, the linear predictor specification for the cross-sectional model in each year is (we suppress time indices here)

$$\tilde{\eta}_i = \tilde{x}_i' \tilde{\beta} + \tilde{v}_s^{\text{stratum}} + q_i' \tilde{v}_r^{\text{AZW}} + \tilde{v}_{ba}^{\text{age}} + \tilde{w}_{rb}\,, \tag{15}$$

where $\tilde{\beta}$ is the fixed effects vector associated with covariates $\tilde{x}$ according to the specification

$$\begin{aligned} & sex * (ageclass6 + background) + sizeclass3 * ageclass3 + \\ & AZW * (sex + nonwestern + income + hhtype + \\ & \qquad sizeclass3 + contract) + \\ & Oversampled + Urban + Province\,. \end{aligned} \tag{16}$$

Further, regarding the random effects,

1. $\tilde{v}_s^{\text{stratum}} \overset{\text{iid}}{\sim} N(0, (\tilde{\sigma}^{\text{stratum}})^2)$ are random intercepts by stratum
2. q_i is the indicator for belonging to the AZW sector or not and $\tilde{v}_r^{\text{AZW}} \overset{\text{iid}}{\sim} N\left(0, \begin{pmatrix} (\tilde{\sigma}_1^{\text{AZW}})^2 & 0 \\ 0 & (\tilde{\sigma}_2^{\text{AZW}})^2 \end{pmatrix}\right)$ are the corresponding 2-vector effects for each

RegioPlus region r.

3. $\tilde{v}_{ba}^{\text{age}} \overset{\text{iid}}{\sim} N(0, (\tilde{\sigma}^{\text{age}})^2)$ denote random intercepts by AZW branch and $ageclass3$.
4. $\tilde{w}_{rb} \overset{\text{iid}}{\sim} N(0, \tilde{\sigma}_w^2)$ are random intercepts by RegioPlus and AZW branch.

For practical reasons, model fitting and prediction for the cross-sectional models are done in the same way as for the time-series models. In particular, the cross-sectional model was fit to all years of NEA data in one pass, by fully interacting each model term with year, allowing for different variance components and residual variances for each year, with independent priors. This way the model is fit like a time-series model, but all parameters and population estimates are effectively those of the cross-sectional model as applied to each year separately. One advantage of this way of fitting the cross-sectional models is that overall model comparison measures like DIC and WAIC can be readily compared to those of the time-series multilevel model.

3.9 Computing small area estimates based on the estimated models

For each of the eight target variables we wish to estimate the corresponding indicator for each domain in the extensive set listed in Table 2.1. This defines a large set of small area estimands. For each target variable the selected model is estimated using MCMC simulation, resulting in a set of draws from the posterior distribution for all the model's parameters. Using these posterior draws, we can subsequently simulate from the posterior predictive distributions for the small area estimands.

Let $\theta_{td} \equiv \frac{1}{N_{td}} \sum_{i \in U_{td}} y_{ti}$ denote a specific domain mean of interest. Here d denotes the domain (e.g. a combination of region and branch), U_{td} is the set of all employees in the population of that domain in year t, $N_{td} = |U_{td}|$ is its size, and y is one of the target variables. Every MCMC draw s ($s = 1 \ldots S$) from the posterior distribution of the model parameters yields a draw from the posterior predictive distribution for θ_{td},

$$\theta_{td}^{(s)} = \frac{1}{N_{td}} \left(\sum_{i \in s_{td}} y_{ti} + \sum_{i \in U_{td} \setminus s_{td}} y_{ti}^{(s)} \right), \tag{17}$$

where the first term sums the observed values over the set s_{td} of NEA respondents (excluding item-non-respondents regarding variable y) in domain d and year t, and the second term adds the simulated predictions for all other employees in the population of domain d in the same year. The draws $y_{ti}^{(s)}$ are generated according to the distribution f in (3). For example, in case of the binomial model for binary data,

$$\begin{aligned} y_{ti}^{(s)} &\sim Be\left(p_{ti}^{(s)}\right), \\ p_{ti}^{(s)} &= \text{logit}^{-1}\left(X_{ti}\beta^{(s)} + \sum_{\alpha} Z_{ti}^{(\alpha)} v^{(\alpha)(s)} \right), \end{aligned} \tag{18}$$

with $\beta^{(s)}$ and $v^{(\alpha)(s)}$ corresponding to the sth MCMC draw for the model coefficients. Together, the S draws obtained this way for $\theta_{td}^{(s)}$ form an approximation of its posterior distribution. We use the means of this approximated distribution as point estimates. Standard errors and credible intervals can be easily computed from this distribution as well.

For the multinomial models, predictions are generated using (10), based on the samples $\tilde{p}_{tik}^{(s)}$ for categories $k = 1 \ldots K-1$. The predictions are then aggregated to subpopulation totals

$$\Theta_{tdk}^{(s)} = \sum_{i \in s_{td}} y_{tik} + \sum_{i \in U_{td} \setminus s_{td}} y_{tik}^{(s)}, \tag{19}$$

for $k = 1 \dots K - 1$, where the posterior sample value for the remainder category is easily derived as $\Theta_{tdK}^{(s)} = N_{td} - \sum_{k=1}^{K-1} \Theta_{tdk}^{(s)}$. For all categorical absence indicators considered there is a 'non-eligible' category, not necessarily the last category, and the indicators of interest are defined as the fraction of each eligible category disregarding the non-eligible category. So if k_{ne} denotes the non-eligible category, the posterior samples for the indicators of interest are given by

$$\theta_{tdk}^{(s)} = \frac{\Theta_{tdk}^{(s)}}{\sum_{k' \in \{1 \dots K\} \setminus k_{\mathrm{ne}}} \Theta_{tdk'}^{(s)}}, \tag{20}$$

for $k \in \{1 \dots K\} \setminus k_{\mathrm{ne}}$.

Regarding computational efficiency of the prediction stage, a few remarks are in order:

1. For a given target variable we carry out the prediction part in a loop over years. This way we only need to load the population microdata for one year at a time, avoiding excessive memory usage. For each year the target variable must be predicted for the population minus the relatively small observed part, amounting to prediction for over 7 million employees per year on average.
2. To reduce the number of predictions we aggregate the population minus the observed part with respect to all variables that make up the model, i.e. we count the number of employees for all unique combinations of all model variables, use these counts in the data generation processes (e.g. Bernoulli sampling becomes binomial sampling) so that we only need a single prediction per unique cell. However, since our model contains many different covariates, including detailed ones such as RegioPlus and AZW branch, the number of unique cells is still very large, only a little smaller than half the population size. Nevertheless, this yields a computational speed-up of the prediction part by a factor of about 2. Note that this approach is very similar to the Multilevel Regression and Poststratification (MrP) approach (Gelman et al., 2016; Gao et al., 2021). In general, the poststratification step requires the availability of counts for the full cross-classification table of model variables, and not necessarily access to the full population microdata.
3. To deal with the large number of predictions in a memory-efficient way, the unit-level or poststratification cell predictions for each Monte Carlo sample are immediately aggregated to the output level defined by the list of table specifications in Table 2.1. That is, the aggregation step is carried out within the loop over Monte Carlo samples, so that there is no need to store the detailed predictions (18) for more than a single Monte Carlo iteration at a time.
4. To further reduce computation time, the predictions are computed in parallel using multiple CPU cores, where we parallelize over Monte Carlo samples.

3.10 Benchmarking

For each target indicator, the model-based small area estimates for the set of domains described in Table 2.1 are consistent in the sense that aggregation of the detailed figures exactly yields the higher-level estimates. This is true because all estimates are computed from the same set of Monte Carlo draws, based on the same model. However, at a high level, several figures have already been produced based on the NEA weights and published on Statistics Netherlands' publication site StatLine. NEA observes many more variables related to working conditions than just the sickness absence variables, and the use of a single set of weights helps to maintain consistency among all derived estimates. The model-based estimates do not exactly agree with the official NEA figures at the highest aggregation level. The differences are expected to be small since the NEA design and weighting variables are also included in the SAE models used.

The model-based small area estimates can easily be adjusted so as to be exactly consistent with the official NEA figures at the overall level. This procedure is common in small area estimation and is called benchmarking. We use a procedure that minimizes a weighted sum of squared differences between original and benchmarked small area estimates subject to the single constraint that the aggregate estimate equals the official NEA figure. For practical reasons we carry out this procedure for each year separately, where we benchmark to the single overall mean NEA figure denoted by $\hat{\bar{Y}}_t$ for year t. Then, if $\hat{\theta}_t$ denotes the ($M = 718$)-dimensional vector corresponding to the complete set of small area mean estimates for year t, the vector of benchmarked small area estimates $\hat{\theta}^{(b)}$ is found by minimizing

$$(\hat{\theta}_t^{(b)} - \hat{\theta}_t)' V_t^{-1} (\hat{\theta}_t^{(b)} - \hat{\theta}_t) \tag{21}$$

subject to $R'\hat{\theta}_t^{(b)} = \hat{\bar{Y}}_t$, where R is the M-vector with value 1 in the first position corresponding to the overall population mean and values 0 elsewhere, and V_t is the (MCMC estimate of the) $M \times M$ posterior covariance matrix for the complete set of small area means for year t. The solution is

$$\hat{\theta}_t^{(b)} = \hat{\theta}_t + V_t R (R' V_t R)^{-1} (\hat{\bar{Y}}_t - R'\hat{\theta}_t)\,. \tag{22}$$

As V_t encodes all consistency relations among the model-based estimates, the benchmarked small area estimates are still internally consistent while also being consistent with the survey-weighted estimate of the overall mean $\hat{\bar{Y}}_t$. As this is done for each year, the set of benchmarked small area estimates is consistent with the series of previously published official NEA figures at the overall level. The benchmarking procedure for categorical indicators uses an expanded version of (22), so that the model-based estimates for each of the $K - 1$ categories are benchmarked simultaneously to the corresponding overall survey-weighted estimates. Finally, we note that (22) does not guarantee that the benchmarked estimates are always positive. However, all benchmarked estimates turned out to be positive.

4 Results

In this section we compare the small area estimates based on the time-seres and cross-sectional models with direct estimates based on the NEA survey weights, for all 8 absence indicators. Comparisons are shown in selected time-series plots and scatter plots, as well as in tables with model criteria and performance measures. Whereas the model criteria are used only for comparing the time-series and cross-sectional models, the performance measures are used to compare the time-series and cross-sectional model estimates to the direct estimates and to each other.

We consider three performance measures, where in the definitions below $\hat{\theta}_{td}^{M}$ denote either time-series or cross-sectional model-based estimates for year t and domain d, and $\hat{\bar{Y}}_{td}$ are direct estimates. Furthermore $v(\hat{\bar{Y}}_{td})$ denote variance estimates of the direct estimates and se($\hat{\theta}_{td}^{M}$) denote the standard errors corresponding to the model-based estimates, i.e. the (Monte Carlo approximations of the) posterior standard errors for the domain means of interest. The three measures are

1. MRB: Mean relative bias as a percentage of direct estimates averaged over time:

$$\mathrm{MRB}_d = 100 \frac{\sum_t (\hat{\theta}_{td}^M - \hat{\hat{Y}}_{td})}{\sum_t (\hat{\hat{Y}}_{td})} \tag{23}$$

2. MRBse: Mean relative bias as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains

$$\mathrm{MRBse}_d = 100 \frac{\sum_t (\hat{\theta}_{td}^M - \hat{\hat{Y}}_{td})}{\sqrt{\sum_t v(\hat{\hat{Y}}_{td})}} \tag{24}$$

3. RRSE: Mean relative reduction of direct standard error averaged over time, by group of publication domains

$$\mathrm{RRSE}_d = 100 \frac{\sum_t \left(\sqrt{v(\hat{\hat{Y}}_{td})} - \mathrm{se}(\theta_{td}^M) \right)}{\sum_t \sqrt{v(\hat{\hat{Y}}_{td})}} . \tag{25}$$

These performance measures are consequently summarized by group of publication domains, as defined in the last column of Table 2.1. Thus, the tables with performance measures display the group-wise minimum, 1st quartile, median, mean, 3rd quartile and maximum of the measures MRB_d, MRBse_d and RRSE_d.

4.1 Results for percentage absence time

The first target variable studied is percentage absence time, i.e. the total number of absent days over the last twelve months divided by the total number of workable days. This is percentage data. We divide by 100 so that the range is between 0 and 1. More than 50% of the response values are 0 (in the 2022 NEA data, it is somewhat less than 50%), and a few hundred response values per year equal 1.

In a previous study (Boonstra et al., 2021) several models have been fitted to the NEA data for a single year (2019). For the percentage absence both linear multilevel and two-part models were considered. The two-part models take into account the zero-inflated character of the data, by modelling the binary zero-indicator variable using a Bernoulli multilevel model and the positive response values using a linear multilevel model. The results were not very different from those based on a univariate linear multilevel model, which is one of the reasons that we here opt for the simpler linear model. The other reason is that the time-series extension would make the two-part models even more computationally intensive.

The estimates based on the time-series multilevel models are compared to the direct estimates as well as to model-based estimates based on the cross-sectional multilevel model estimates for each year of NEA data. In Table 4.1 the DIC and WAIC criteria for the fitted time-series and cross-sectional models are given. Lower values indicate a better trade-off between model fit and model complexity. The DIC/WAIC values for the time-series model are about 700 units lower than those for the cross-sectional model fits combined over all years, implying a strong preference for the time-series model. Note also that the estimated effective number of model parameters according to both criteria (columns p_DIC and p_WAIC) are much higher for the cross-sectional models combined over all years.

Figure 4.1 compares the time-series and cross-sectional model-based estimates to the direct estimates, at the aggregate levels of the total population of employees, and the subpopulation of AZW employees. For the total population the time-series and cross-sectional model-based estimates are nearly equal, which is also true of their respective 95% intervals, which almost completely overlap. The model-based estimates and intervals are also not much different from the direct estimates and intervals, except that the direct estimates are slightly higher, especially in the years before 2018. The difference is small in an absolute sense, though not completely negligible given the relatively small standard errors at these high levels of aggregation. The difference in levels might be caused by a different correction for non-response bias due to the survey weighting as compared to the multilevel modelling and prediction approach, e.g. due to differences in the exact set of covariates/weighting variables used over time. This conjecture is partially confirmed because the covariates $income$ and $contract$ were added to the weighting model only in 2022 and it was observed that they generally yield a stronger non-response correction compared to the previously used educational covariate. Indeed, Figure 4.1 shows that the model-based estimates closely match the overall-level direct estimate in 2022. The fact that $income$ and $contract$ are used as covariates in the multilevel models over the entire period 2014-2022 may then, at least partially, explain the differences with respect to the direct overall estimate in earlier years.

To ensure that the model based estimates agree exactly with the direct estimates at the highest aggregation level, the time-series multilevel model estimates are benchmarked to the direct estimates, see the purple dotted line in Figure 4.1 (left). For the AZW subpopulation (Figure 4.1, right) there are already some differences between time-series and cross-sectional model estimates, where the latter are somewhat more variable over time, with slightly wider confidence bands. The estimated standard errors of the direct estimates are higher than those based on both models. At this level the benchmarked estimates based on the time-series model are still slightly higher than the unbenchmarked estimates, though no longer equal to the direct estimates. Note the different scales on the y-axes; the absence percentage is clearly higher for the AZW subpopulation.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	-505631	385	-505454	562
cross-sectional model	-504932	1102	-504742	1293

Table 4.1 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the percentage sick leave NEA data of 2014-2022.

Figure 4.2 contains similar time-series plots for a selection of four AZW branches. At these more detailed levels we already see much narrower uncertainty intervals for the model-based estimates compared to the direct estimates, especially for relatively small AZW branches such as GPs and health centers and Youth care. The time-series model estimates display somewhat smaller standard errors than the estimates based on the cross-sectional models. Also noticeable is the difference in mean level over time between time-series and cross-sectional model estimates for Youth care and, to a lesser extent, UMCs. Here the average level of the time-series model estimates is much closer

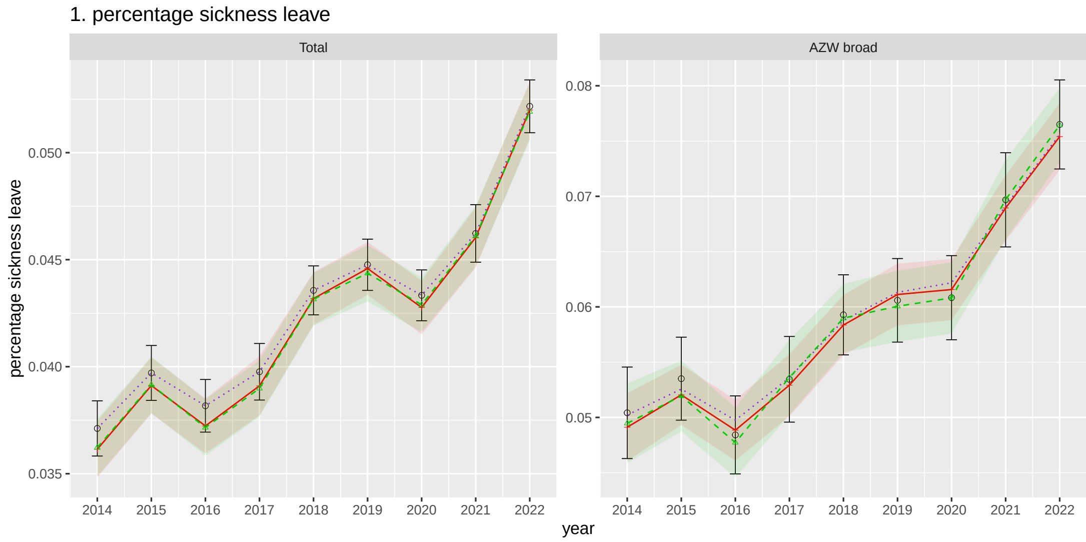

Figure 4.1 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right). In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

to the average level of the direct estimates. This is a big advantage of borrowing strength over time. The cross-sectional estimates are shrunk towards a regression mean level independently in each year, resulting in a systematic underestimation for small domains with relatively high absence percentages, such as Youth care, and systematic overestimation for small domains with relatively low absence percentages such as UMCs. Modelling over time largely corrects such biases. Another expected difference is that the time-series model estimates exhibit smoother dependence over time with narrower uncertainty intervals compared to the cross-sectional and direct estimates.

Figure 4.3 shows time-series plots for Region Noord-Nederland, Province Gelderland and RegioPlus Zuid-Holland Zuid, both for the overall absence percentage and that for AZW narrow, i.e. the AZW population minus the childcare branch. Here we see some indications that the cross-sectional models are overfitting: especially for Noord-Nederland and Gelderland the cross-sectional estimates seem quite close to the direct estimates despite quite large jumps over time and large standard errors of the driect estimates. This is probably due to the fact that Province has been included as fixed effects in the cross-sectional multilevel model. By including them as random effects instead this type of overfitting might be reduced. The time-series model estimates do not seem to overfit the data in this sense. The time-series model contains fixed effects for the more detailed RegioPlus classification, but in that case the coefficient estimates are pooled over time. Time dependence is instead modelled using random walk effects.

For the most detailed estimation level Figure 4.4 shows some example time-series plots for the combination of two RegioPlus regions (Drenthe and Rijnmond) and two AZW branches (Hospitals and other MCs, and Disability care). At this detailed level the number of observations are generally very small, which is why the direct estimates have such large standard errors. The model-based estimates are also more uncertain at this level of detail, even though the model-based standard errors are much smaller compared to the direct standard errors. It is also clear that the standard errors for the time-series model estimates are somewhat smaller than those for the cross-sectional model estimates.

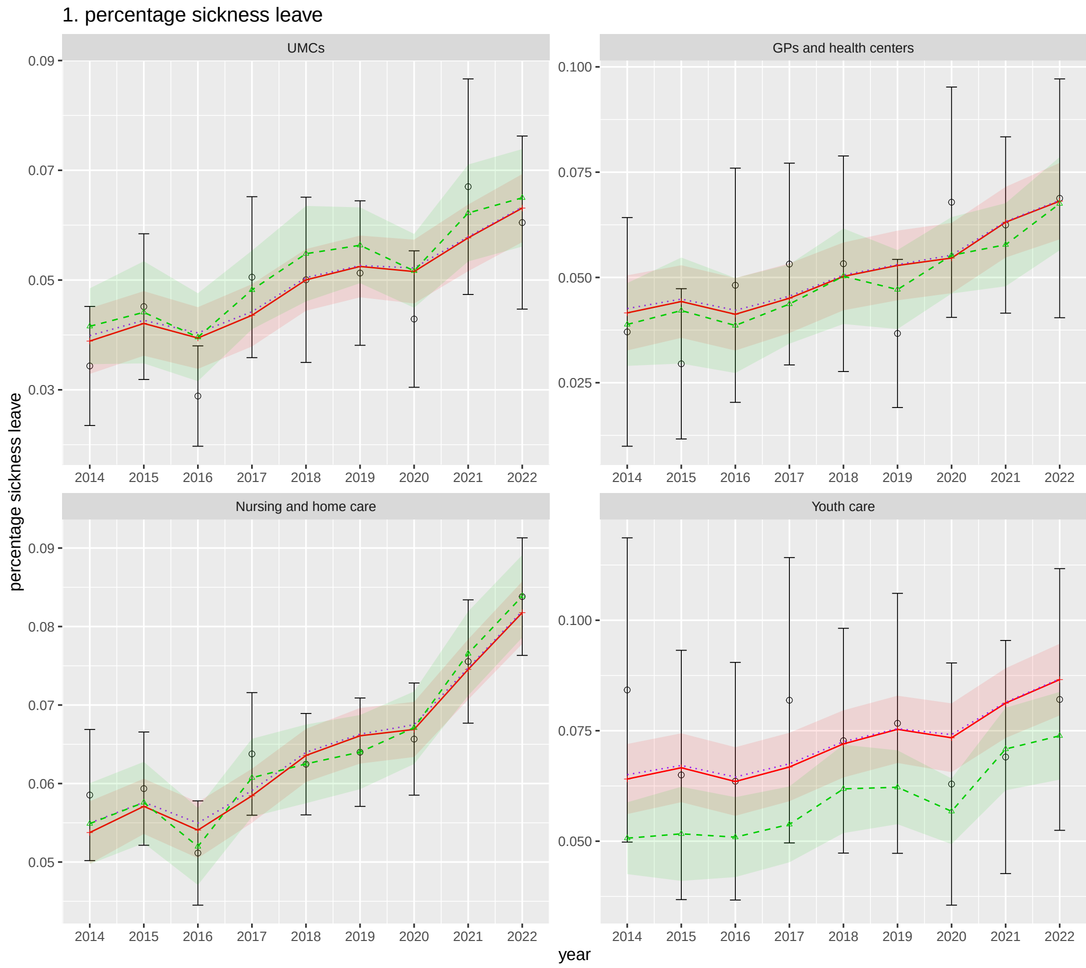

Figure 4.2 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

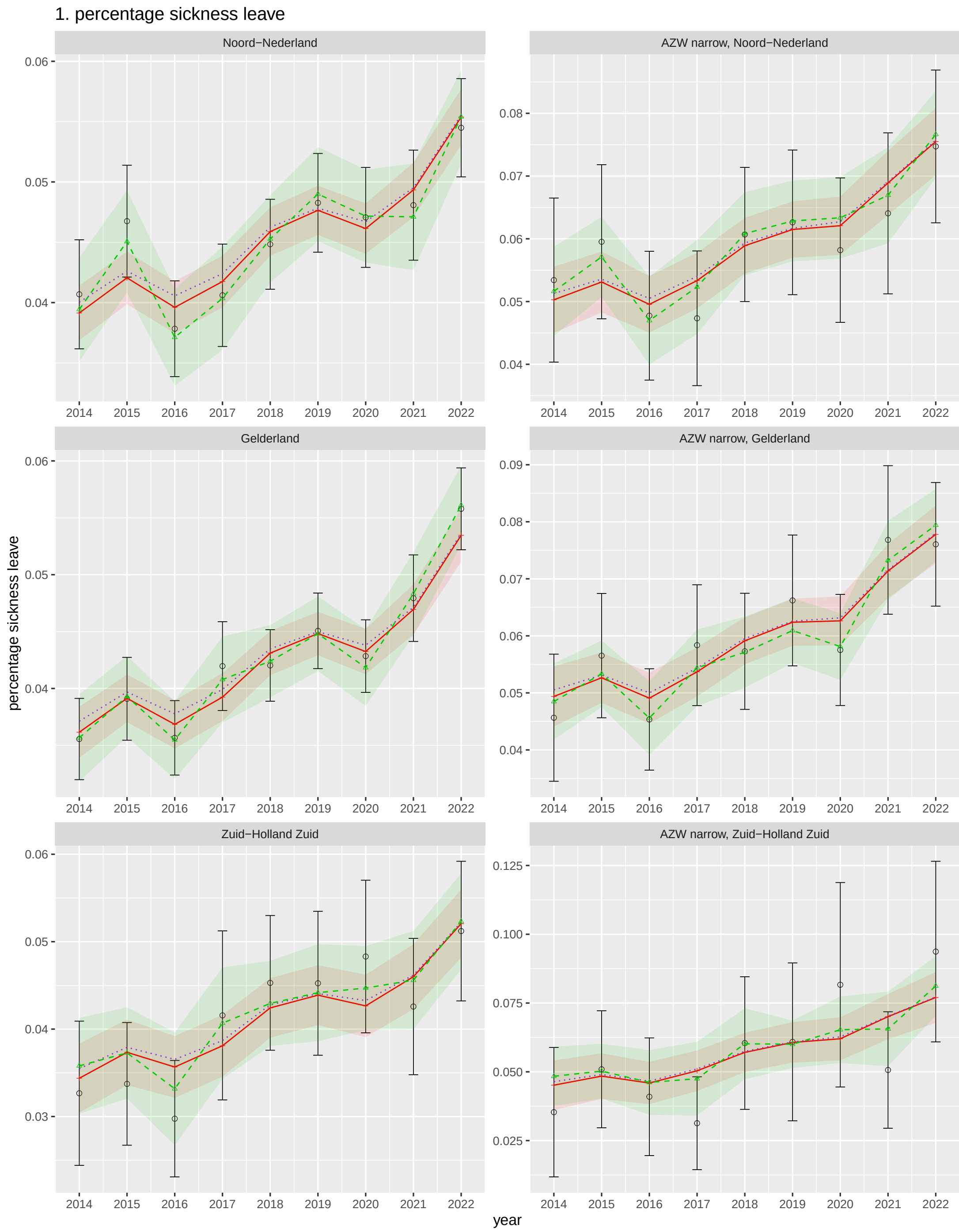

Figure 4.3 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

For more time-series plots, for a selection of RegioPlus and Age class by AZW branch domains, see Appendix C.1.

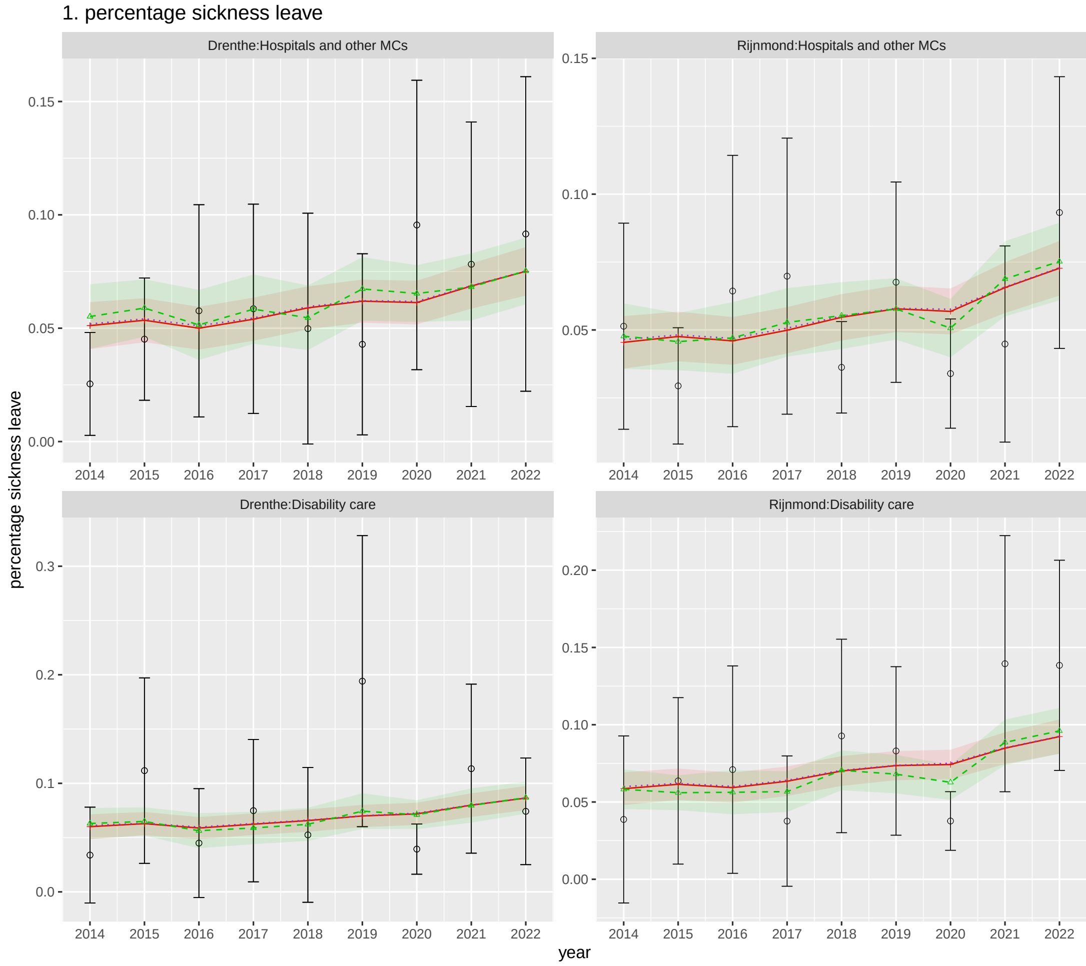

Figure 4.4 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

Figure 4.5 shows scatter plots of model-based against direct estimates, for all 718 domain estimates for year 2022. The leftmost plot shows the huge difference in range of the direct point estimates versus the model-based estimates. The model-based estimates are practically all between 0.03 and 0.10, whereas there are many direct estimates that lie far beyond these values and some of them even equal 0 or 1. Note that larger domains are generally closer to the diagonal black line, whereas extreme direct estimates correspond to very small domains. The middle panel clearly shows the large reduction in standard error of the model-based estimates compared to the direct estimates. It also shows that the time-series model yields mostly somewhat smaller standard errors as compared to the cross-sectional model. This can also be seen in the rightmost plot, a scatter plot of the coefficients of variation (CVs), i.e. relative standard errors. Note that also here the dots and triangles for larger domains are generally closer to the diagonal. A CV of 0.25 was chosen as a limit above which estimates are not published. For this indicator we see that all model-based estimates shown in Figure 4.5 are below this limit.

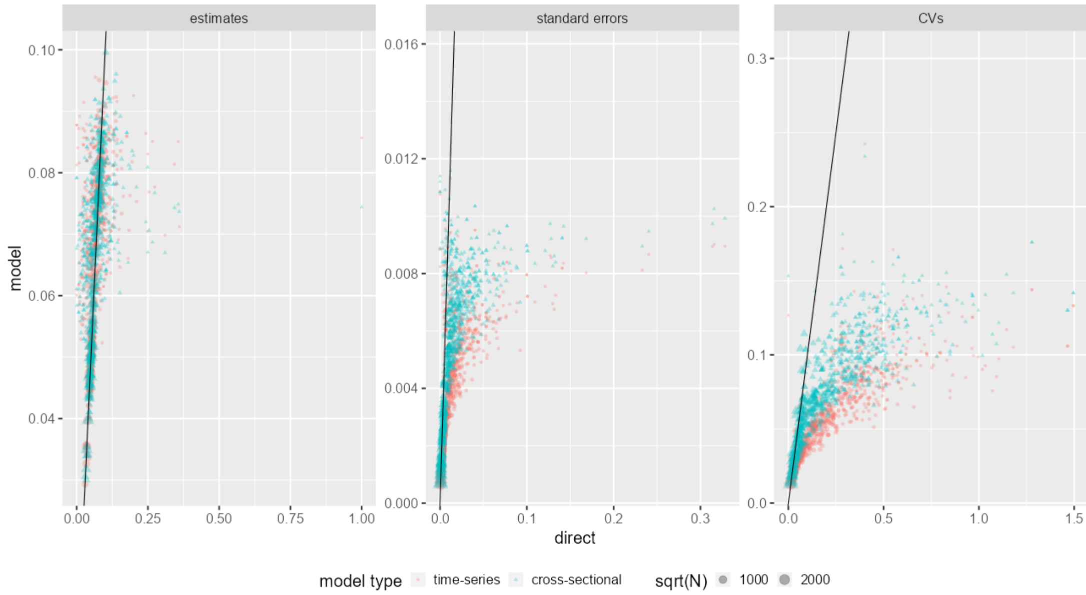

Figure 4.5 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for percentage absence in 2022. The red dots correspond to the time-series model and the blue dots to the cross-sectional model.

Tables 4.2-4.4 contain summaries of the MRB, MRBse and RRSE measures as defined in equations (23)-(25), by group of publication domains, where the groups are defined in (the last column of) Table 2.1. From Table 4.2 we see that the mean of the model-based estimates over time can deviate substantially from that of the direct estimates. By far the largest deviations appear in the detailed domains by region and branch. This is no surprise as for many of those domains the direct estimates are very unreliable, even when averaged over all years in the range 2014-2022. But even at high aggregation levels there are still negative relative deviations of over 15%. Relative to the standard errors of the direct estimates, these differences are even larger, see Table 4.3. We have seen this already in Figure 4.1 at the overall level, the differences being largest in the years before 2018. However, these differences at the overall level are removed by benchmarking, so that the final benchmarked time-series model estimates are in exact agreement with the direct figures at this level. Tables 4.2 and 4.3 also show that averaged over time the time-series model estimates are in general closer to the direct estimates than are the cross-sectional model estimates, a phenomenon also already observed in some of the time-series plots, see e.g. Figure 4.2. Finally, Table 4.4 shows that the standard errors of the model-based estimates are generally much smaller than those of the direct estimates, especially for the more detailed domains. Also, the time-series model estimates have smaller standard errors than the cross-sectional model estimates, confirming the depiction in Figure 4.5. Note that the large negative entries in Table 4.4 for the most detailed RegioPlus-branch domains are due to unreliable direct standard error estimates, which are sometimes estimated at 0, e.g. when all observations in a small domain happen to equal 0.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-16.5	-1.5	-0.8	-1.6	-0.4	1.8
	cross-sectional	-16.6	-1.7	-0.9	-1.6	-0.3	1.8
AZW branch	time-series	-13.8	-2.6	-1.2	-0.4	1.0	11.0
	cross-sectional	-26.2	-5.9	-1.0	-2.1	3.6	15.0
region	time-series	-2.8	-1.8	-1.1	-0.6	-0.1	4.1
	cross-sectional	-5.1	-2.4	-1.5	-0.8	-0.1	6.4
RegioPlus	time-series	-7.5	-2.2	-0.9	-0.5	0.5	10.0
	cross-sectional	-14.9	-3.0	-1.0	-0.1	2.1	22.9
region-branch	time-series	-44.9	-7.4	-0.2	5.8	9.2	289.6
	cross-sectional	-44.4	-9.5	-1.2	3.8	9.8	257.4
RegioPlus-branch	time-series	-67.9	-8.8	0.5	9.8	15.5	289.6
	cross-sectional	-72.1	-11.7	0.2	8.8	18.8	257.4

Table 4.2 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-133.0	-43.8	-17.4	-26.7	-8.0	47.3
	cross-sectional	-133.5	-46.8	-18.6	-27.8	-6.5	46.6
AZW branch	time-series	-82.4	-16.1	-5.0	-8.1	4.5	59.1
	cross-sectional	-96.1	-37.0	-10.7	-11.3	17.8	78.7
region	time-series	-68.3	-19.6	-10.5	-12.1	-1.2	30.2
	cross-sectional	-67.5	-37.0	-14.2	-14.3	-2.1	55.5
RegioPlus	time-series	-48.9	-19.3	-6.5	-7.1	5.4	32.6
	cross-sectional	-95.8	-23.6	-6.4	-6.0	16.1	69.0
region-branch	time-series	-76.3	-22.5	-2.2	1.6	21.8	148.6
	cross-sectional	-90.6	-29.4	-5.1	-0.6	22.3	162.1
RegioPlus-branch	time-series	-90.4	-23.0	0.0	7.6	21.5	544.2
	cross-sectional	-91.8	-26.8	-2.3	7.6	29.6	546.0

Table 4.3 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-18.8	20.8	36.4	31.5	46.5	53.7
	cross-sectional	-69.9	-4.5	8.6	4.5	19.2	39.5
AZW branch	time-series	-29.8	56.6	65.6	60.1	71.1	80.4
	cross-sectional	-55.1	36.5	51.7	46.4	61.6	77.6
region	time-series	13.8	49.6	56.9	54.1	61.9	72.7
	cross-sectional	-6.0	3.9	36.7	27.2	41.5	52.3
RegioPlus	time-series	42.0	55.4	62.7	61.0	67.5	72.7
	cross-sectional	-1.7	28.0	43.2	38.2	49.8	59.3
region-branch	time-series	11.0	70.7	77.6	74.1	81.9	89.9
	cross-sectional	-7.2	59.3	69.5	63.4	76.5	88.1
RegioPlus-branch	time-series	-118.0	71.9	77.9	74.4	82.9	97.1
	cross-sectional	-130.6	63.2	71.4	65.9	78.6	96.9

Table 4.4 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

4.2 Results for percentage of employees with absence

The next indicator is the percentage of employees who have been absent due to sickness in the past twelve months, corresponding to a binary target variable at the employee level. This target variable has fewer item non-responses than the other target variables. Here the choice of sampling distribution is straightforward, as a Bernoulli sampling distribution is most appropriate, and combined with a logistic link function the corresponding multilevel models (6) can still be estimated efficiently. For this particular target variable, a linear (Gaussian) model would also be appropriate as the percentages of interest are far away from 0 and 1, and we have found that small area estimates and their standard errors based on the Bernoulli and linear multilevel models are indeed nearly the same. The final estimates presented in this section are based on the Bernoulli multilevel time-series model.

Table 4.5 contains the estimated model criteria for the time-series and cross-sectional binomial multilevel models (combined over all years). The time-series model is favoured over the cross-sectional model, having about 700 units lower DIC and WAIC criteria values.

Figure 4.6 shows time-series plots for the overall percentage sickness leave as well as that for the AZW broad subpopulation. It immediately stands out that the estimates for 2022 are much higher than those of the preceding years (though note that the y-axis scales do not start at 0). This jump is most likely largely due to a redesign of the NEA questionnaire in 2022. We also note that the model-based estimates again have a negative bias relative to the direct estimates, especially at the overall level and for the years before 2018. This deviation is probably due to a different non-response correction for the model-based estimates as compared to that of the direct estimates. The deviation at the overall level is overcome by benchmarking the time-series model estimates, as shown by the purple dotted line. The time-series and cross-sectional model estimates and standard errors are quite similar at this aggregation level and their dependence over time is very similar to that of the direct estimates.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	607781	458	607781	457
cross-sectional model	608435	1084	608437	1083

Table 4.5 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the percentage of employees with absence NEA data of 2014-2022.

Figure 4.7 shows that at the AZW branch level the model-based estimates appear to be more accurate than the direct estimates. Standard errors of the time-series model estimates are slightly smaller than for the cross-sectional model, and several differences in mean level over time are visible, where the time-series model estimates are on average closer to the direct estimates' average over time. Figure 4.8 shows time-series plots for a selection of several RegioPlus domains. Here we see some indications that the cross-sectional model is overfitting the data, especially for Flevoland. This may be related to the fact that Flevoland is both a province and a RegioPlus region, and that the cross-sectional model contains fixed effects for all Provinces. Further time-series plots as well as tables with performance measures can be found in Appendix C.2.

Figure 4.9 shows scatter plots of model-based versus direct estimates. Note the smaller range of the model-based estimates compared to the direct estimates (left), and the much smaller standard errors and CVs of the model-based estimates. The time-series

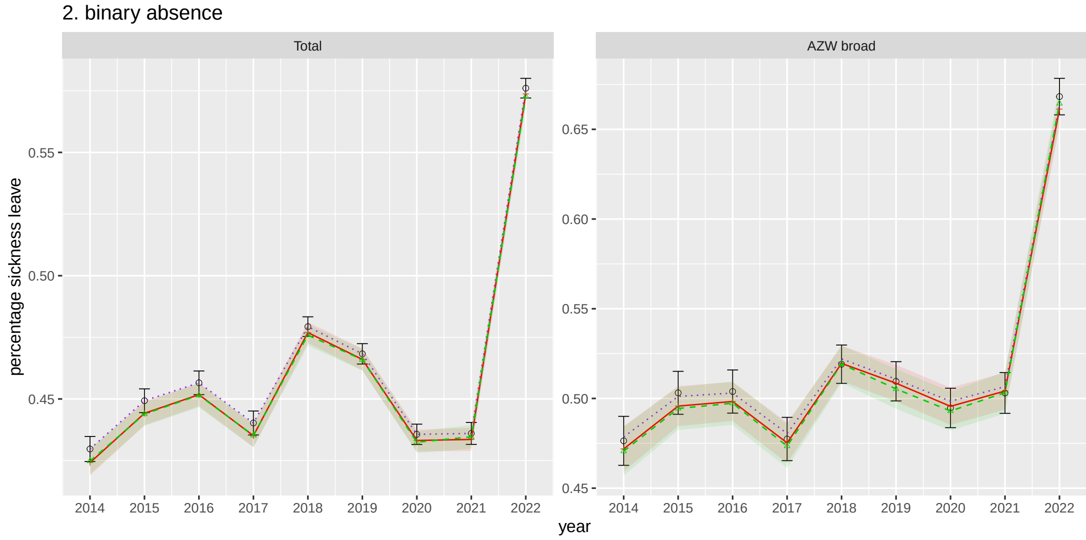

Figure 4.6 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right). In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

model estimates mostly have somewhat smaller standard errors than the cross-sectional model estimates. It is also worthwhile to note that the model-based CVs for this indicator are quite a bit smaller than those for the other indicators.

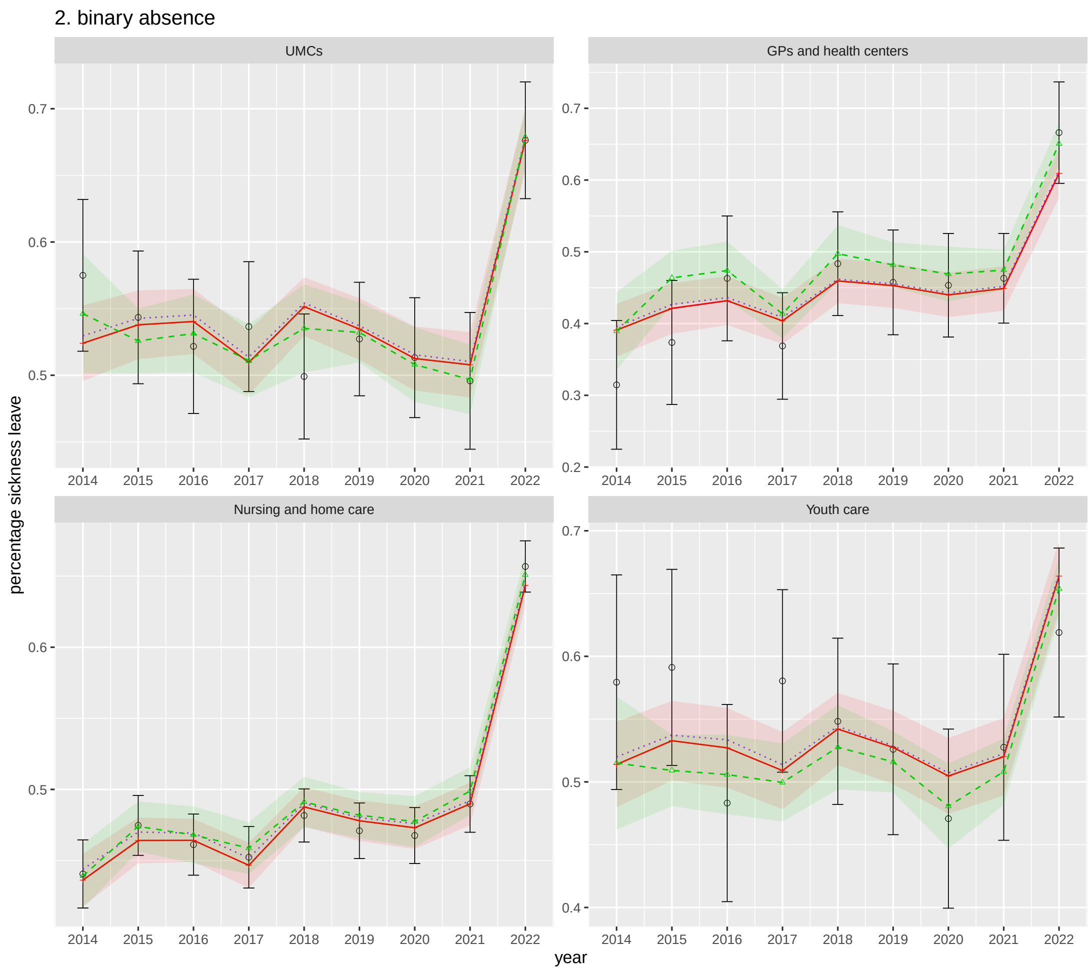

Figure 4.7 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

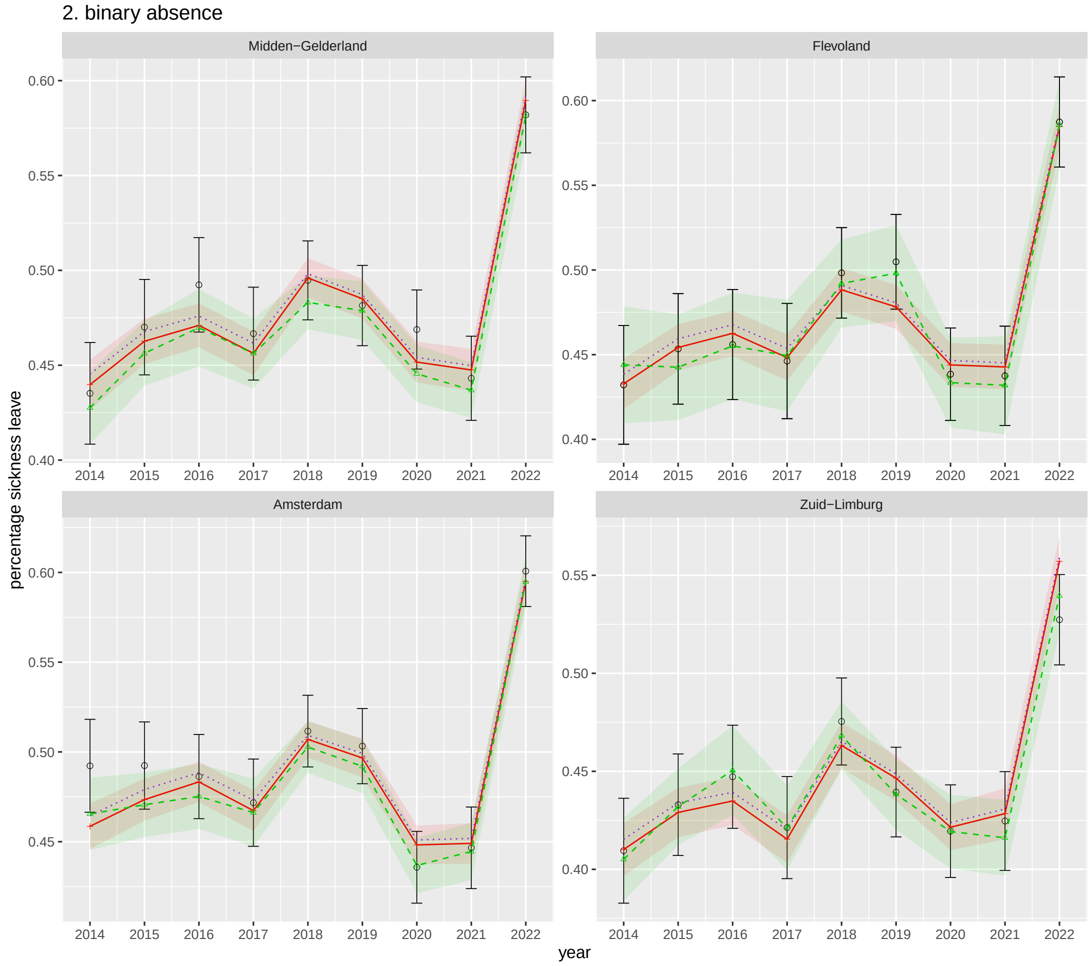

Figure 4.8 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

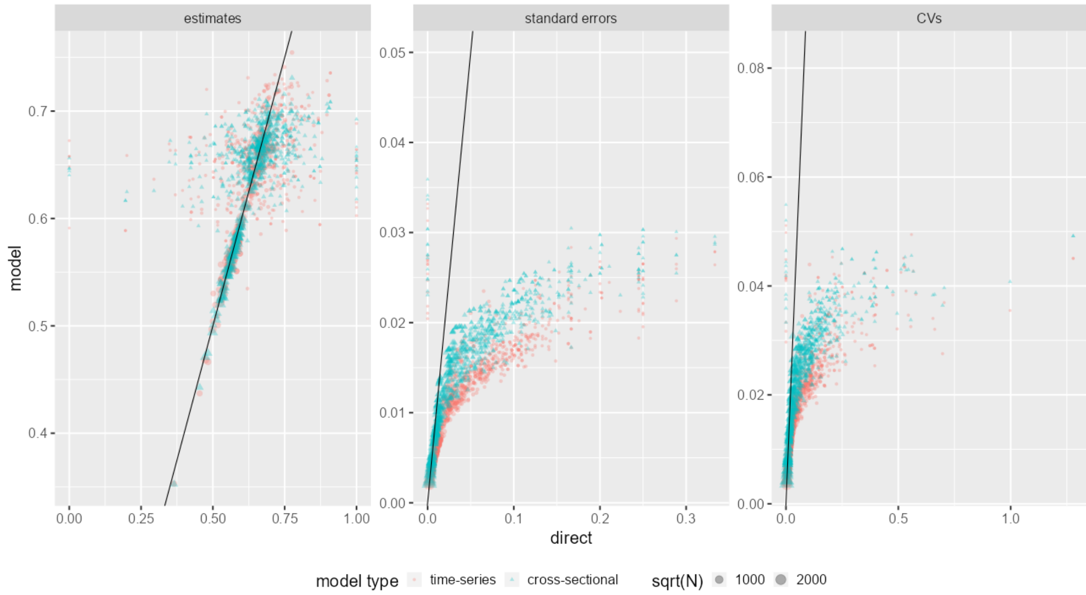

Figure 4.9 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for the percentage of employees with absence in 2022.

4.3 Results for number of absence periods

The third target variable is the number of absent periods due to sickness over the last twelve months. This is a count variable with a skewed distribution with over 50% zero observations in each year except 2022, where the percentage is quite a bit lower at 42%, probably due to the questionnaire redesign. The maximum possible value is 40, which is attained by about 100-300 respondents per year.

We first tried to fit a negative binomial multilevel model to this data. However, the negative binomial model estimates showed signs of overfitting, as well as non-robustness against the large maximum values of 40 in several smaller domains. Instead, it turned out that a linear time-series multilevel model resulted in more plausible estimates. In addition, the linear time-series multilevel model produced much better posterior predictive p-values for a set of test statistics, including the overall variation of data and time-variation at several aggregation levels.

The selected linear time-series multilevel model fit is compared to the corresponding combined set of linear cross-sectional multilevel model fits in Table 4.6. Again, the DIC and WAIC model criteria show a clear preference for the time-series model, whose values are roughly 700 units lower.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	2356430	336	2356862	767
cross-sectional model	2357074	1039	2357547	1515

Table 4.6 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the number of absence periods NEA data of 2014-2022.

Figure 4.10 shows that for this indicator the differences between (unbenchmarked) model and direct estimates are small at the overall level, except perhaps in 2014. At the overall and AZW broad levels the differences between time-series and cross-sectional model estimates are also quite small. Note the relatively high values in 2022, coinciding with the NEA questionnaire redesign.

More differences can be seen at the level of individual AZW branches, see Figure 4.11. Here the time-series model estimates are much more smooth compared to the direct estimates and slightly smoother than the cross-sectional estimates. There are some differences in mean level over time between the time-series and cross-sectional model estimates, where the former is generally closer to the mean level of the direct estimates. Note also that the standard errors of the model estimates, especially the time-series ones, are smaller than those of the direct estimates. Further time-series plots as well as tables with performance measures can be found in Appendix C.3.

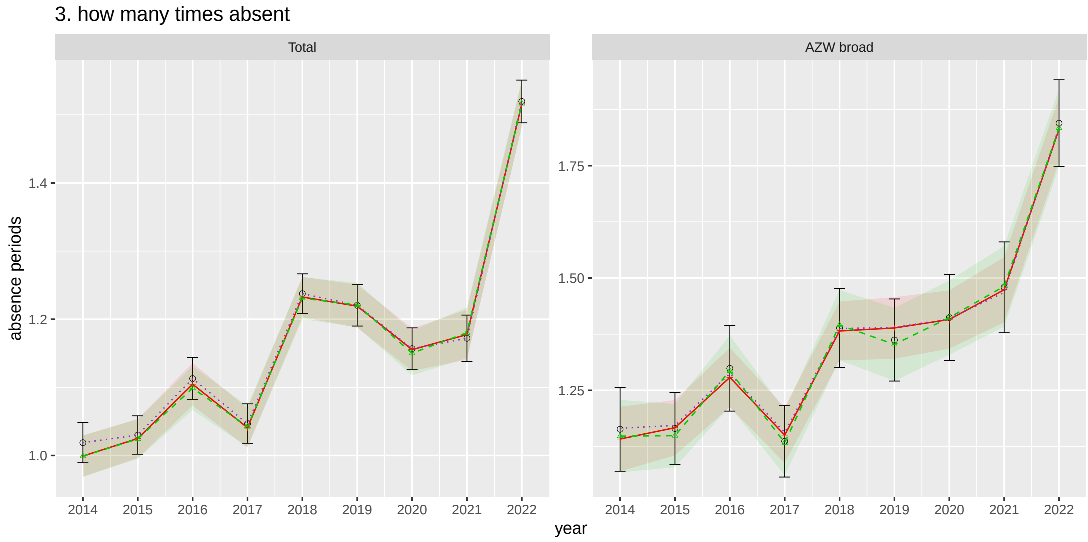

Figure 4.10 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right). In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

Figure 4.12 shows, for the subset of estimates for 2022, the smaller range of the model estimates as well as their smaller standard errors compared to the direct estimates. The time-series model estimates' standard errors are mostly somewhat smaller than those of the cross-sectional model, and for both subsets of estimates all CVs are below 25%.

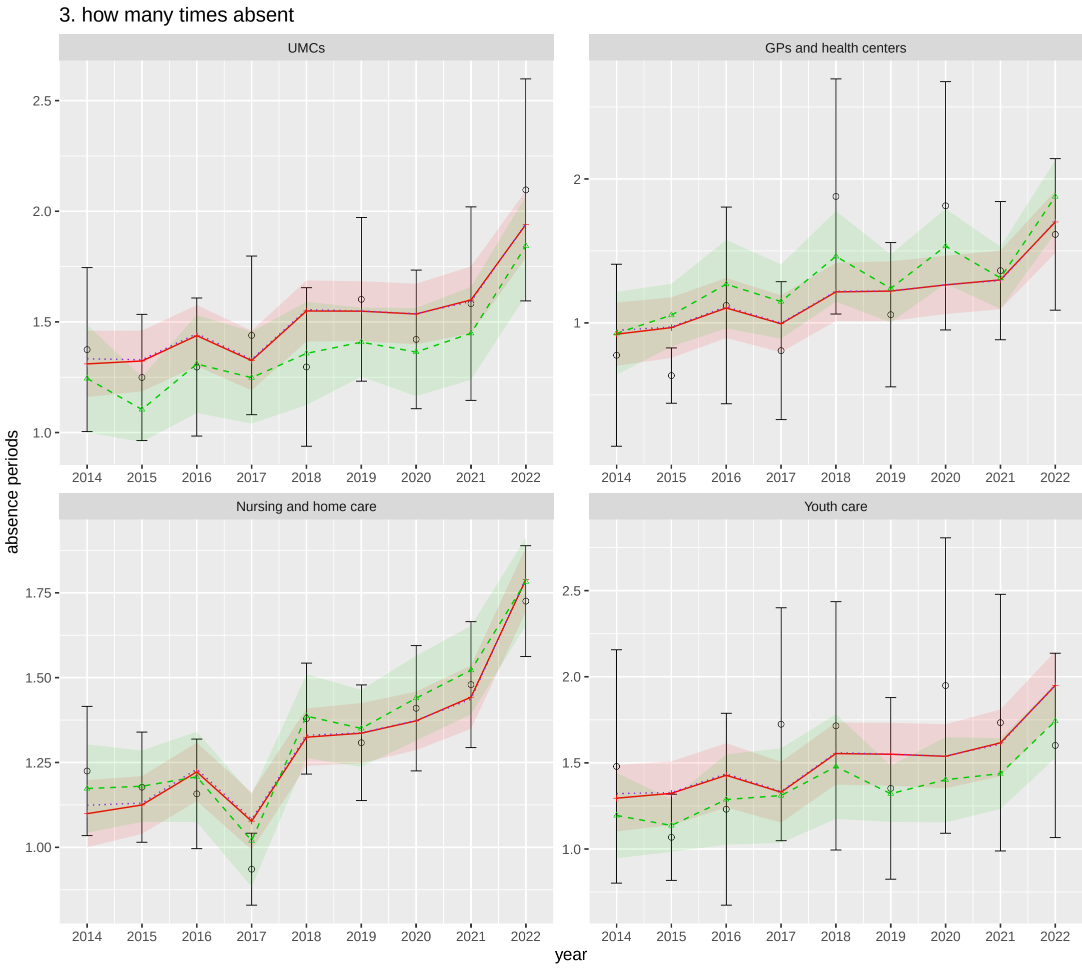

Figure 4.11 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

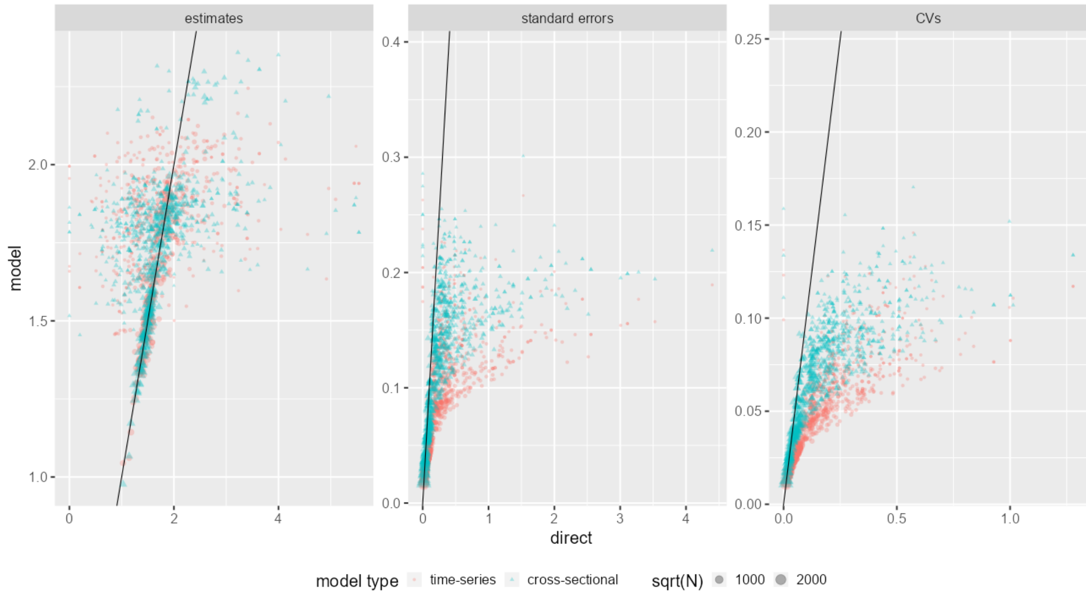

Figure 4.12 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for number of absence periods in 2022.

4.4 Results for the number of absence days

The next target variable is the actual number of absent days during the last 12 months. Like the number of absence periods, this variable is a count variable, but with a wider distribution, varying from 0 to a maximum of 215.

We found similar but smaller differences as in Subsection 4.3 between the results based on a negative binomial multilevel model compared to the corresponding linear multilevel model, and in this case it was harder to say which model yields the overall best results. In the end we selected the negative binomial model as it appears more appropriate for the count data at hand.

Table 4.7 shows the model criteria values found for the estimated negative binomial time-series and (combined) cross-sectional models. The criteria show a preference for the time-series model, with a difference of about 700 units for WAIC and a smaller difference of about 300 units regarding DIC.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	2017882	617	2018279	982
cross-sectional model	2018188	1215	2018983	1898

Table 4.7 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the number of absence days NEA data of 2014-2022.

The time-series plots in Figure 4.13 show that the model-based estimates, especially the time-series ones, are somewhat below the overall direct estimates, with the largest differences in the years before 2018. These differences are similar to those observed for the related percentage and binary absence variables described in Subsections 4.1 and 4.2, and most likely due to a slightly different correction for non-response by the NEA weighting versus the model-based prediction. After the benchmarking step, the time-series model estimates agree exactly with the direct estimates at this overall level, as shown by the purple dotted line in the left panel of the figure.

Figure 4.14 shows the same time-series plots for a selection of AZW branches. Here it can be observed again how sometimes the average level of the time-series model estimates is in closer agreement with that of the direct estimates than is the average level of the cross-sectional model estimates.

Figure 4.13 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right). In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

From Figure 4.15, showing scatter plots of model-based versus direct estimates for all domains for the year 2022, the much smaller, more realistic, range of the model-based point estimates is clear from the left panel. The estimated standard errors of the model-based estimates are much smaller than the direct standard errors. For this indicator, we do not find much difference in standard errors for the 2022 time-series and cross-sectional model estimates. A closer look at the time-series plots (here and in Appendix C.4), and the RRSE performance measures (Table C.9 in Appendix C.4), reveals that generally the time-series model estimates still have somewhat lower estimated standard errors than the cross-sectional estimates, but that this is hardly the case for the last year 2022.

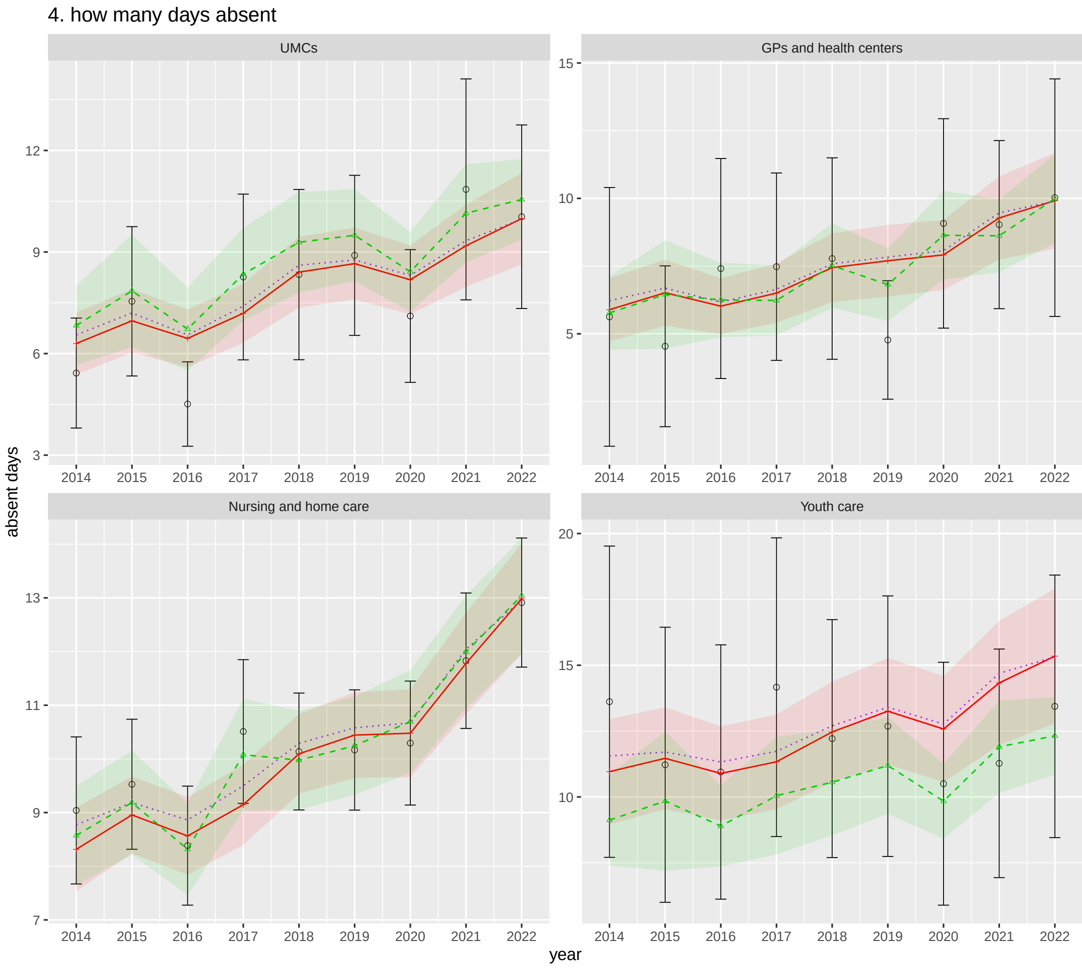

Figure 4.14 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

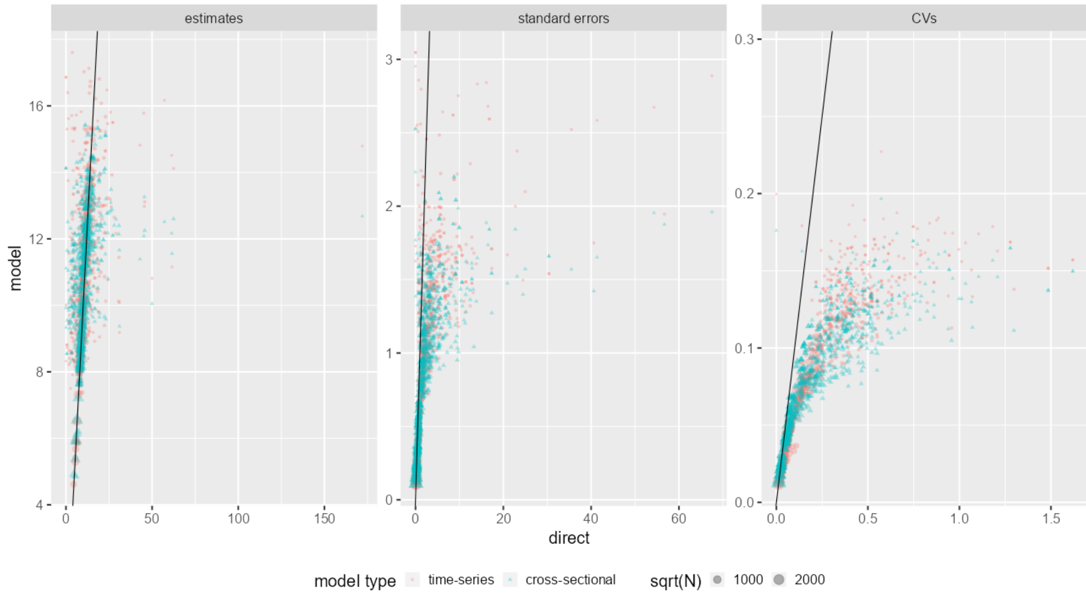

Figure 4.15 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for the number of absence days in 2022.

4.5 Results for the duration of the last absence

The duration of the last absence is defined as a categorical variable with classes 1-4 days, 5-19 days, 20-209 days, 210 or more days, as well as the non-applicable (n/a) category of employees who have never been absent due to sickness. Note that this indicator as well as the remaining ones in the next 3 subsections refer to the last absence, which not necessarily took place in the preceding twelve months.

We fit this target variable, as well as those in the next three subsections, using the time-series and cross-sectional multilevel models with multinomial sampling distribution, see Subsection 3.5. A disadvantage of the computationally efficient stick-breaking representation of the multinomial model is that results may depend on the ordering of the categories. We have verified that in our case this dependence is rather small. In addition, we compared the model information criteria for several orderings of categories and in most cases found somewhat better values if we order the categories from small to large. For the absence duration, however, we found better model criteria for the natural ordering of categories mentioned above, so we did not change it.

From the model criteria values in Table 4.8 it can be concluded that the time-series multilevel model is again preferred over the cross-sectional multilevel model with a difference of about 2500 units.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	985025	1410	985040	1418
cross-sectional model	987430	3984	987624	4097

Table 4.8 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the duration of last absence NEA data of 2014-2022.

Figure 4.16 shows time-series plots for the Total and AZW broad populations for each category. Note that the estimates for the first four (non-n/a) categories have been computed according to (20), and therefore add up to one. The estimates for the non-applicable category are presented here for completeness, but they are not published. The time-series and cross-sectional estimates are quite similar at these high levels of aggregation. Both model-based series follow the direct estimates rather closely, although some systematic differences in level are present. The largest of these systematic differences arises for the non-applicable category, where the model estimates are systematically larger at the overall level. Also observe from Figure 4.16 that the benchmarked time-series model estimates exactly agree with the direct estimates at the overall level.

This response variable also shows some large jumps from 2021 to 2022, most likely related to the NEA redesign. Especially the large downward jump of the n/a category is noteworthy, as this amounts to a reduction of more than 50% in both the overall and AZW broad employee populations. It appears that the non-applicable category, corresponding to respondents who claim to have never been absent due to illness, is most sensitive for biases due to non-response and measurement error.

Figure 4.17 shows time-series plots for the first category, 1-4 work days, for a selection of AZW branches. To save some space, we present most time-series plots for categorical indicators for a single category only. In this figure we once again observe the phenomenon that the time-series model estimates generally better agree with the direct estimates when averaged over time. Also, at this level already the time-series model estimates are somewhat smoother than the cross-sectional model estimates, and have narrower uncertainty intervals. A selection of further time-series plots at other aggregation levels can be found in Appendix C.5. The same Appendix also contains the tables with performance measures, where the summaries are taken over the estimates from all categories.

In Figure 4.18 scatter plots of direct versus model-based estimates for the year 2022 are displayed, for all categories. In the left part of the plot the clustering corresponding to the different categories is well visible, except that categories 20-209 work days and n/a largely overlap. From the middle and right panels we observe that the model-based estimates generally have much smaller standard errors compared to the direct estimates, and the time-series model estimates for 2022 seem to have slightly smaller standard errors compared to the cross-sectional model estimates. Note that there are quite a few domains for which the time-series estimates of this indicator cannot be published according to the CV < 0.25 rule. These cases mostly concern the relatively rare category '210 and more work days'.

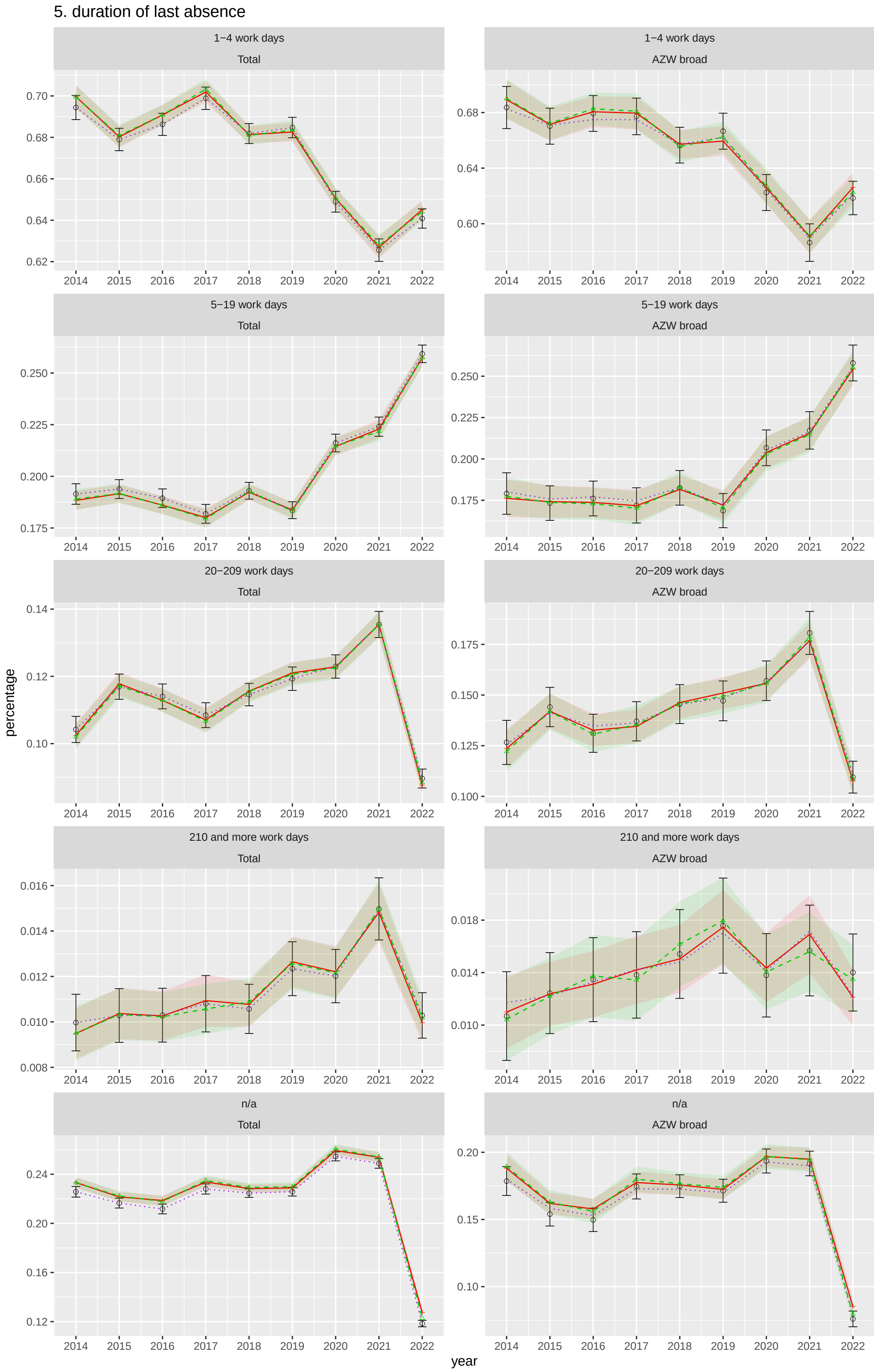

Figure 4.16 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right) for all categories. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

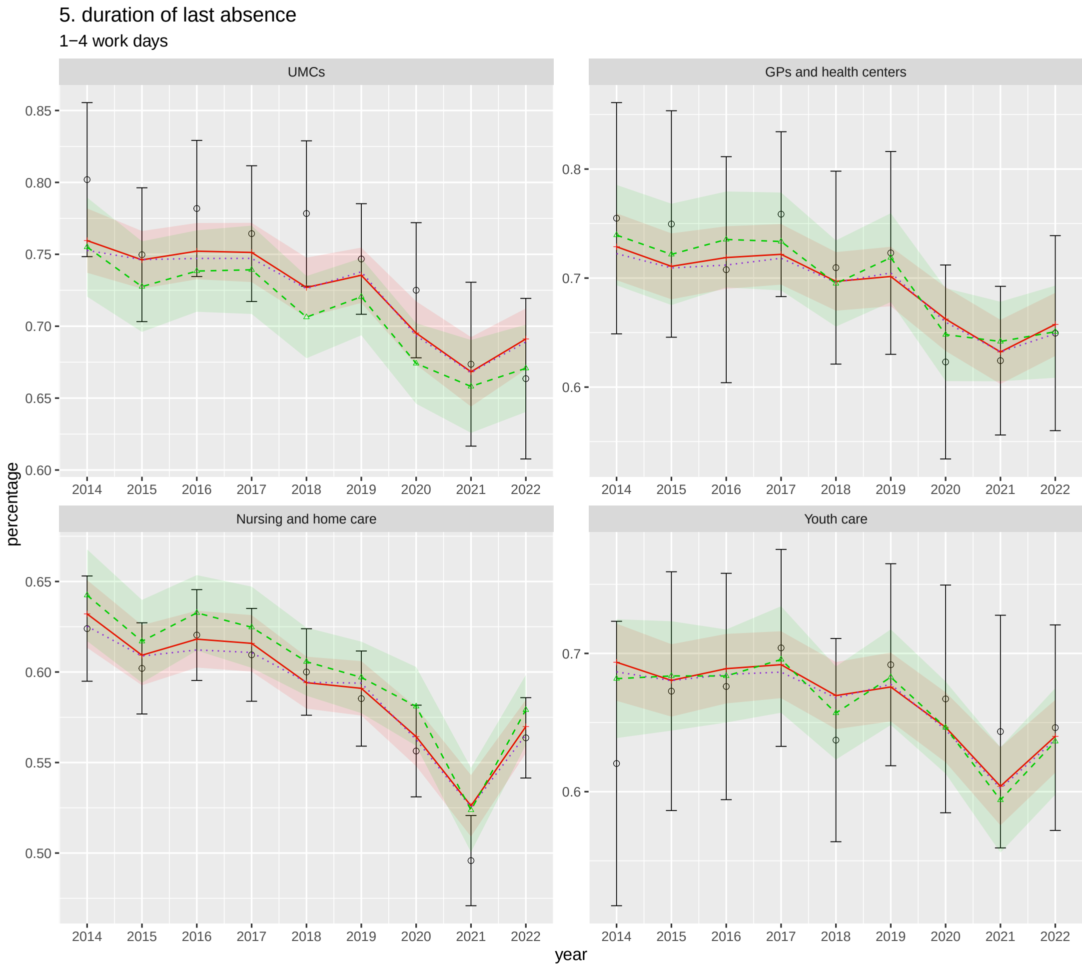

Figure 4.17 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

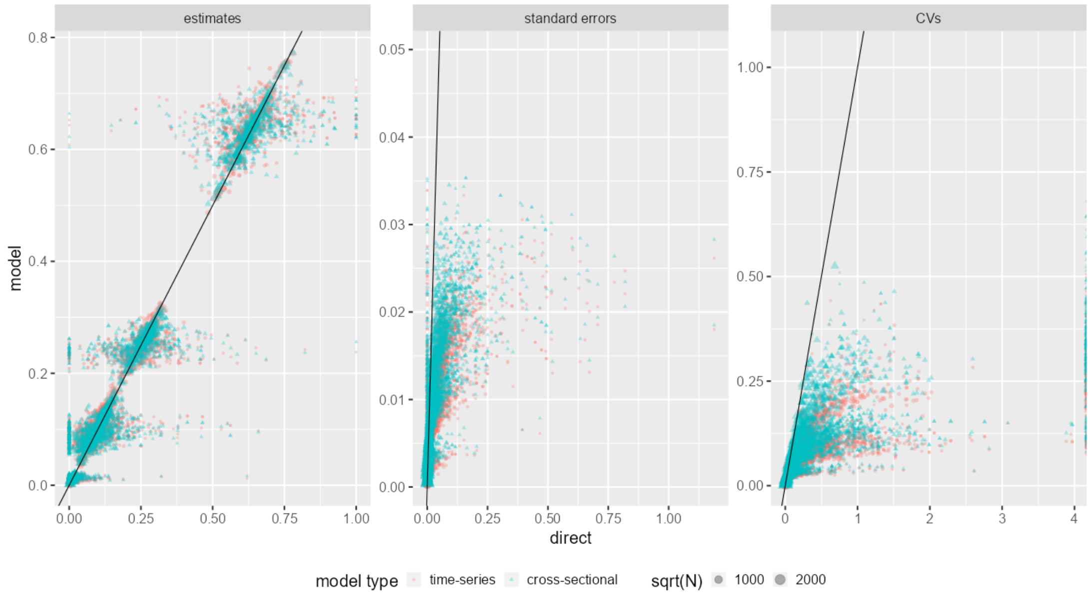

Figure 4.18 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for duration of last absence in 2022.

4.6 Results for the work-relatedness of the last absence

The sixth absence indicator measures the work-relatedness of the last absence, in categories ‘mainly work-related’, ‘partly work-related’, ‘not work-related’, ‘don’t know’ and the not-applicable category corresponding to employees who have never been absent due to sickness. For fitting the multinomial models to this target variable we re-ordered the categories by letting the rarest category ‘don’t know’ to be the first category. This resulted in somewhat better model criteria compared to using the natural questionnaire ordering mentioned above.

The model criteria values in Table 4.9 show a preference for the time-series multilevel model over the cross-sectional multilevel model with a difference of more than 2500 units.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	1034671	1415	1034685	1423
cross-sectional model	1037275	4056	1037353	4100

Table 4.9 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the work-relatedness of last absence NEA data of 2014-2022.

Figure 4.19 shows time-series plots for the Total and AZW broad populations of employees, for all categories. At this level of aggregation, the differences are generally small. The systematic differences between model-based and direct estimates are also smaller for this variable, except for the n/a category where the model estimates are sytematically higher before benchmarking. Note that the non-n/a category estimates have been scaled to sum to 1. In Figure 4.20 the time-series plots are shown for a selection of AZW branches, and for the first category of work-relatedness only. Here we already observe the smaller uncertainty intervals of especially the time-series model estimates compared to the direct estimates. Further time-series plots can be found in

Appendix C.6.

From Figure 4.21 it can be seen that the model-based estimates have a smaller range than the direct estimates, and that the standard errors of the time-series model estimates are smallest for the 2022 estimates. Only for a few estimands the time-series model CV exceeds 0.25. See Appendix C.6 for tables with summaries of the performance measures, combining the estimates for all categories.

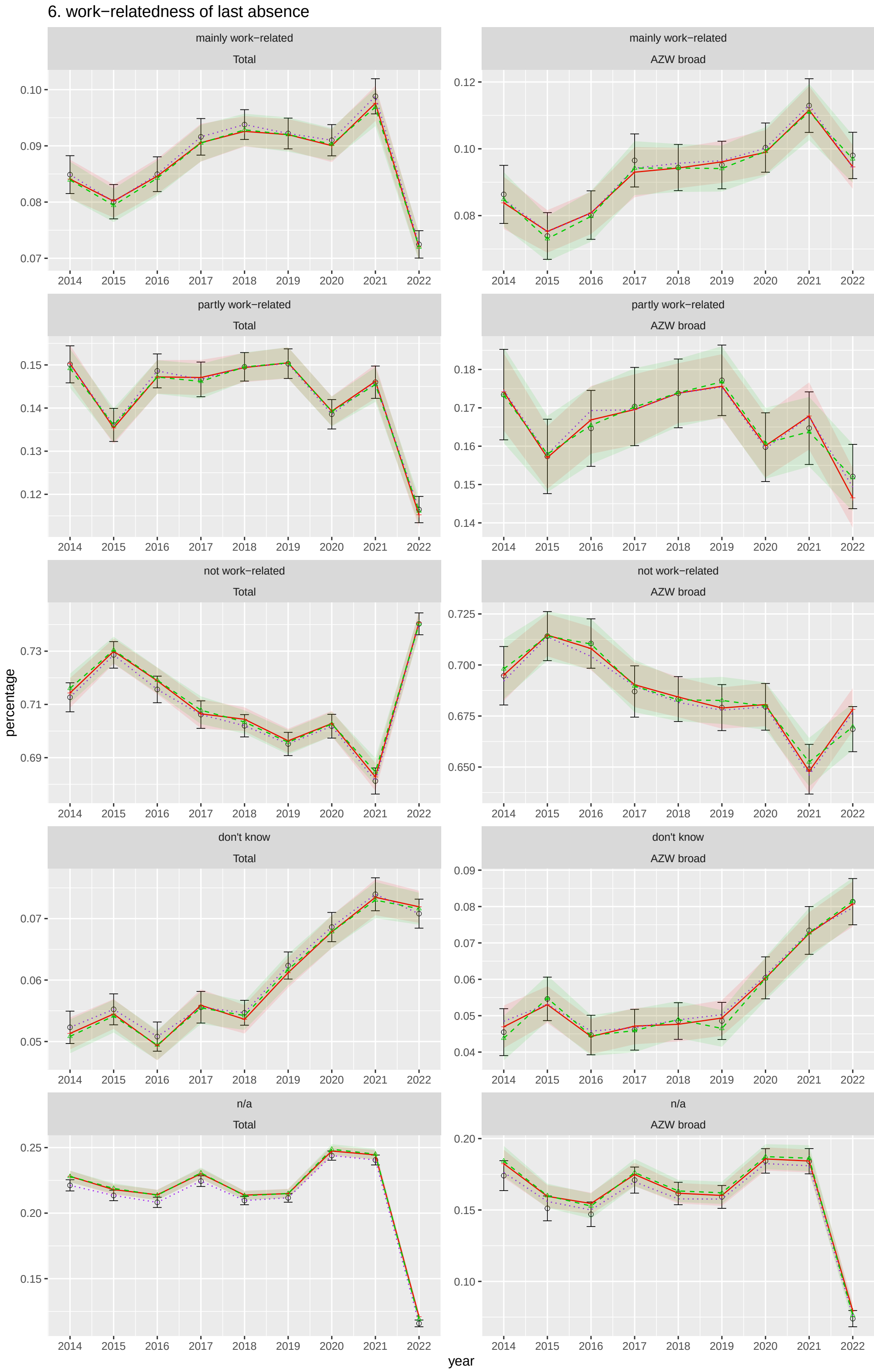

Figure 4.19 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right) for all categories. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

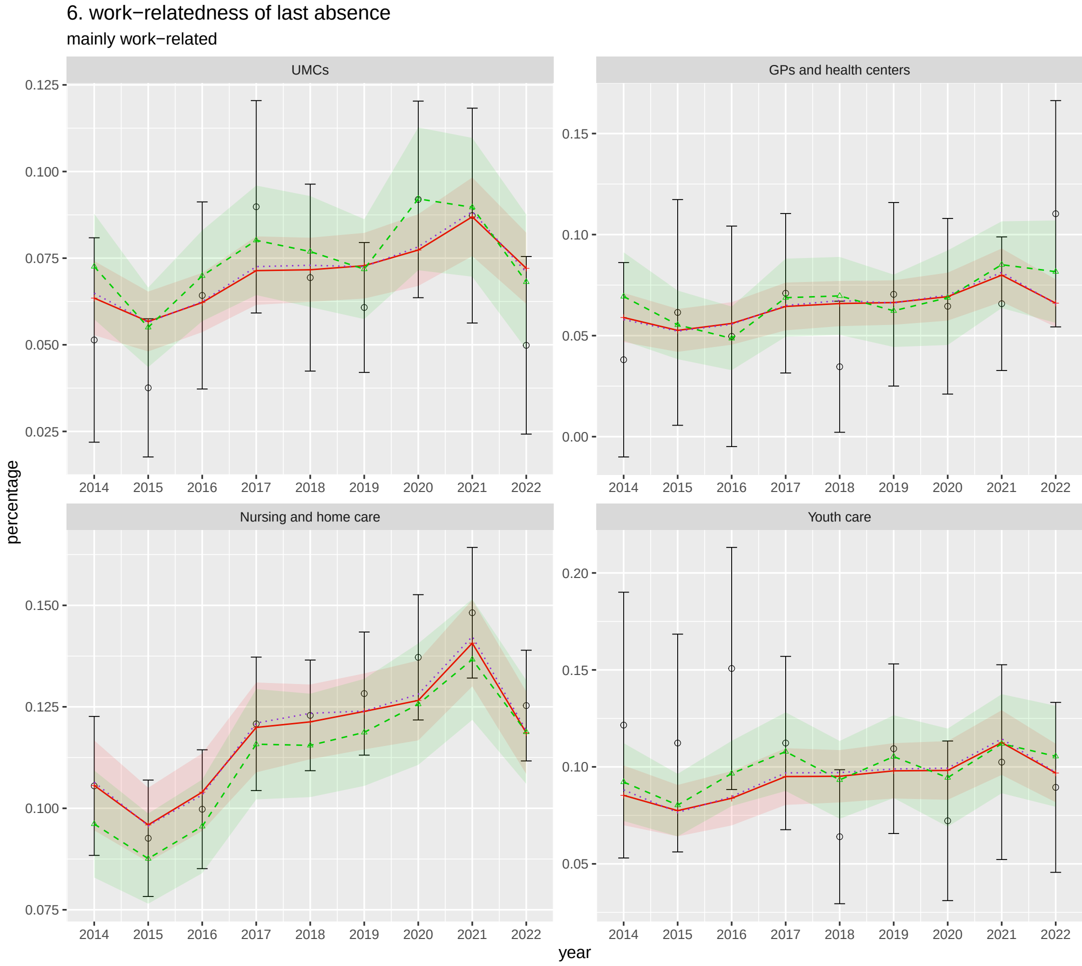

Figure 4.20 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

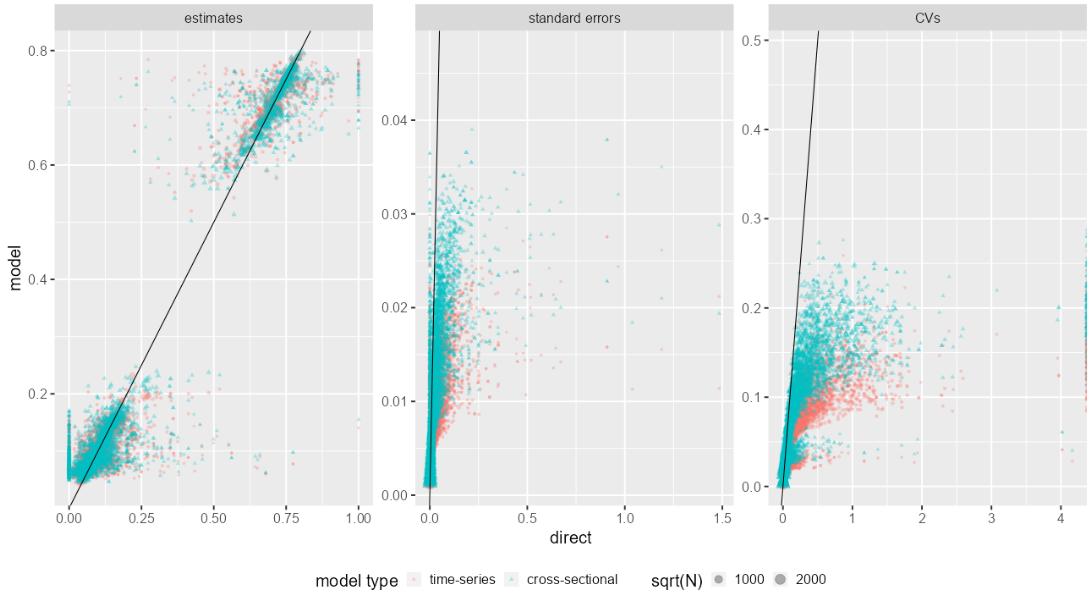

Figure 4.21 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for work-relatedness of the last absence in 2022.

4.7 Results for the type of complaints

For the type of complaints associated with the last absence, the original categories distinguished in the questionnaire are aggregated to the following categories: ‘physical’, ‘psycho-social’ and ‘other’, as well as the non-applicable category corresponding to employees who have never been absent due to sickness. For model fitting the categories have been re-ordered from small to large in terms of numbers of observations, resulting in the order ‘psycho-social’, ‘other’, ‘n/a’ and ‘physical’. This resulted in somewhat better model criteria.

Table 4.10 shows that by the model criteria the time-series model is preferred over the cross-sectional model, by a difference of almost 2000 units.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	870436	974	870439	976
cross-sectional model	872284	2920	872342	2954

Table 4.10 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the type of complaints NEA data of 2014-2022.

Figure 4.22 shows time-series plots for the Total and AZW broad populations of employees, for all categories. Estimates for the first three, eligible categories have been constrained to add to 1. The series show large movements over the last three years. These may be related to the Covid pandemic, but the large jumps from 2021 to 2022 are most probably also to a large extent due to the NEA redesign. In particular, for this target variable, as well as the one described in Subsection 4.8, the original set of underlying questionnaire categories is different in 2022.

Figure 4.23 shows time-series plots for the first category ‘physical’ only, for a selection of 4 AZW branches. Here it can be noted that the time-series model estimates are more smooth and have smaller standard errors compared to both direct and cross-sectional

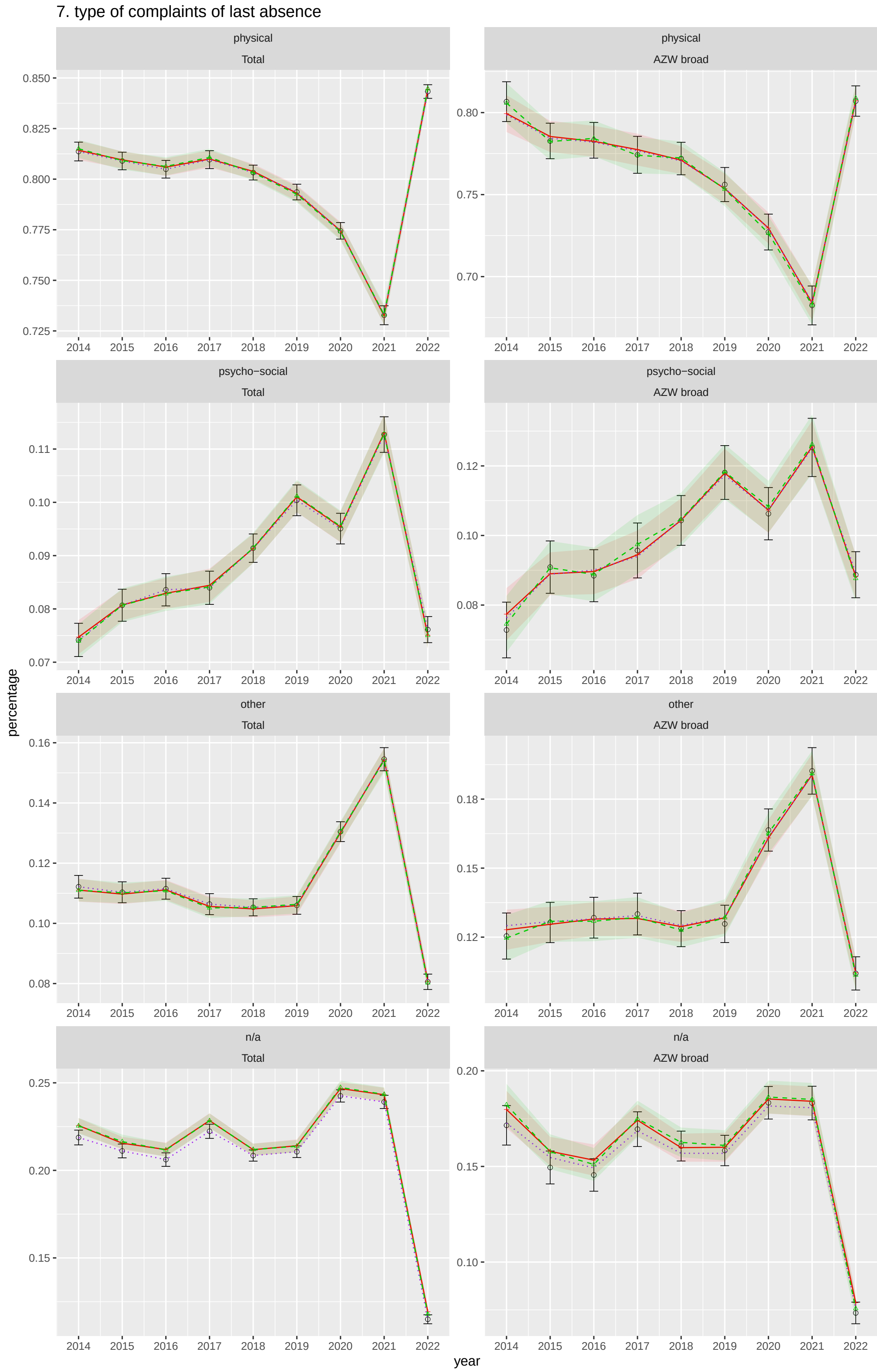

Figure 4.22 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right) for all categories. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

model estimates. Also, on average over the years the time-series model estimates stay closer to the direct estimates. Further time-series plots as well as tables with perfomance measures can be found in Appendix C.7.

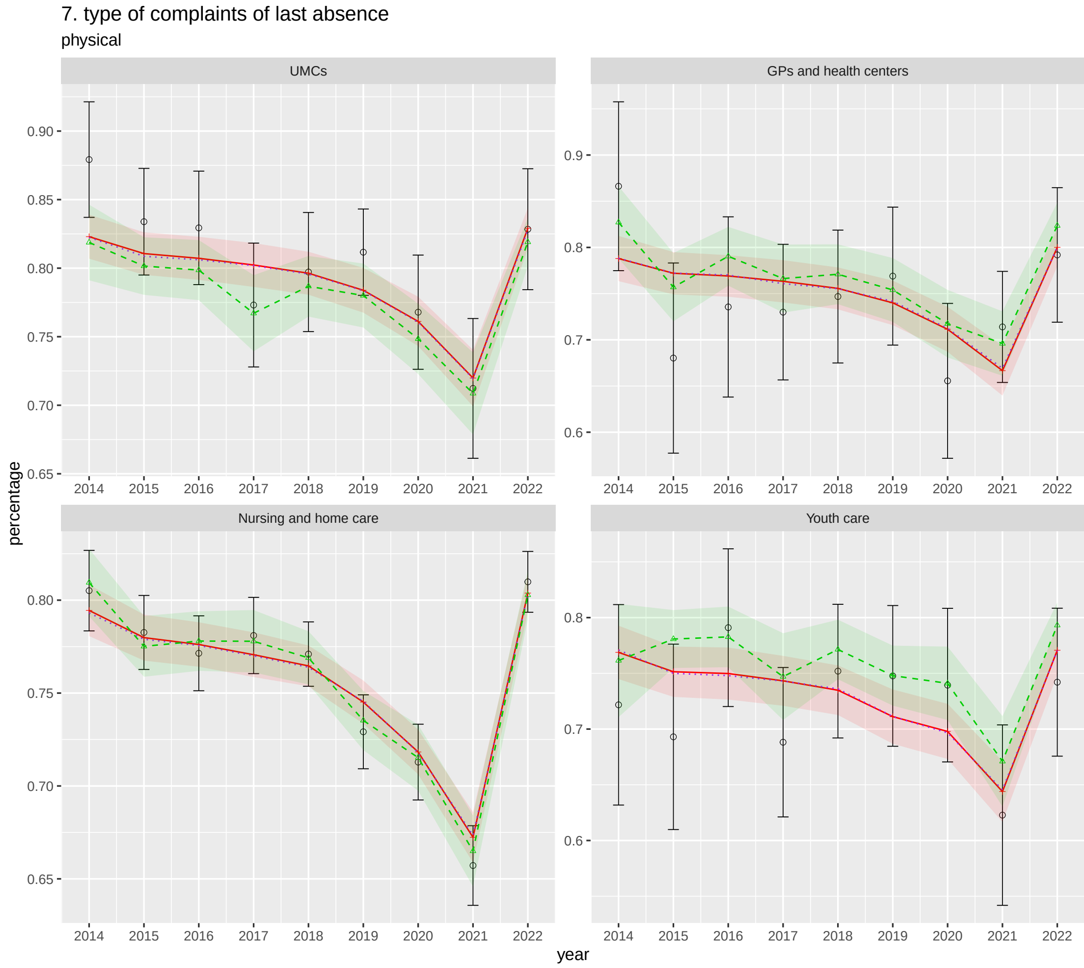

Figure 4.23 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

Figure 4.24 shows scatter plots of model-based versus direct estimates for all categories and all domains of interest for the year 2022. The time-series model estimates (red dots) generally have slightly smaller standard errors compared to the cross-sectional model estimates. For this indicator practically all 2022 time-series model estimates have a coefficient of variation below 0.2.

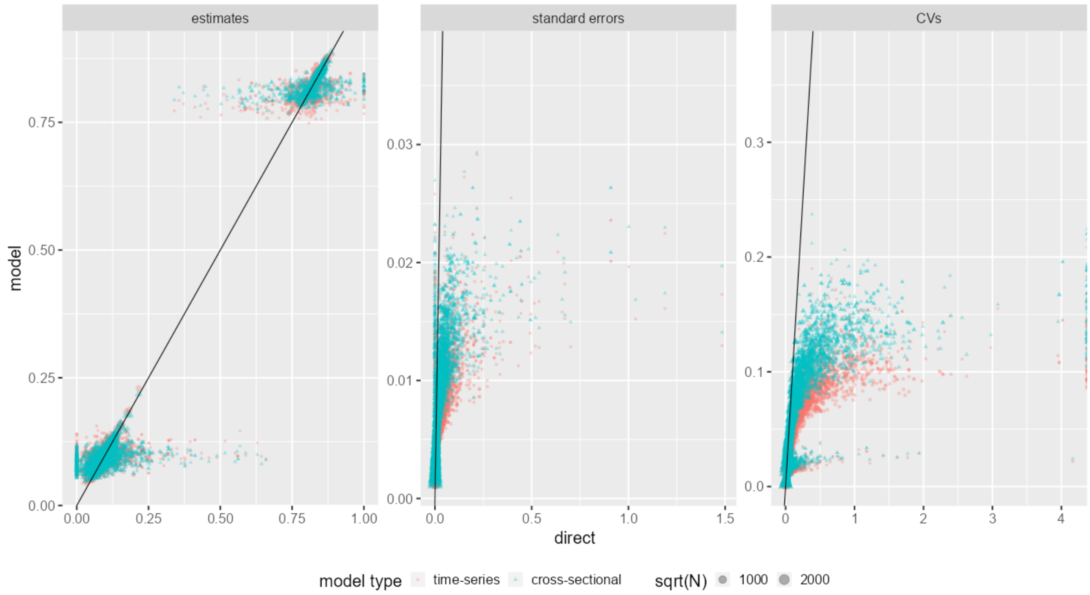

Figure 4.24 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for type of complaints in 2022.

4.8 Results for the most important reason that led to work-related complaints

The 8th and final indicator considered is the most important reason that led to work-related complaints. Here the work-related complaints refer to the last absence due to sickness in case this absence was indeed work-related, i.e. was designated as being 'mainly work-related' or 'partly work-related' by the work-relatedness target variable described in Subsection 4.6. This means that only a relatively small fraction of the population of employees is eligible with regard to this target variable, as the 'n/a' category now includes, besides the employees who have never been absent, all employees whose last absence was not work-related.

Besides 'n/a', the following categories are distinguished: 'physical burden', 'psycho-social burden' and 'other factors'. For fitting the model these have been re-ordered from small to large in terms of numbers of observations, i.e. in the order 'other factors', 'physical burden', 'psycho-social burden' and 'n/a'. The resulting model-criteria for time-series and cross-sectional multinomial time-series multilevel models are given in Table 4.11. The criteria for the time-series model are more than 2000 units lower, so again the time-series model is preferred.

	DIC	p_DIC	WAIC	p_WAIC
time-series model	568437	1001	568446	1006
cross-sectional model	570647	3263	570720	3301

Table 4.11 DIC and WAIC criteria and effective number of degrees of freedom estimates for time-series and cross-sectional multilevel models fitted to the reason of work-related absence NEA data of 2014-2022.

Figure 4.25 shows for all categories the estimates for the Total and AZW broad populations. Note first of all the large fraction of employees in the non-applicable category. The remaining categories have been rescaled to sum to 1. For this indicator the 2022 estimates are again quite different, in particular for the 'psycho-social' and 'other factors' categories. For this indicator the original categories of the underlying questionnaire variable have changed in 2022, and this may largely explain the differences.

From Figure 4.26 we see that at the AZW branch level the time-series model estimates (for category 'physical burden') are indeed an improvement over the cross-sectional model estimates in the sense that they are more smooth, as is justified given the large standard errors of the direct estimates, and on average the level is more compatible with that of the direct estimates, especially for the Youth care branch. Further time-series plots as well as tables with performance measures can be found in Appendix C.8.

The scatterplots of Figure 4.27 refer to the estimates for the last year 2022. Estimates for all categories are combined in these plots. Once again we see from the middle and right panels that the standard errors of the model-based estimates are mostly much smaller than the direct standard errors, and that the time-series model's error estimates and CVs are slightly smaller compared to the cross-sectional model. Note that there are a few time-series model estimates for 2022 with CVs exceeding 0.25, though not nearly as much as cross-sectional model estimates.

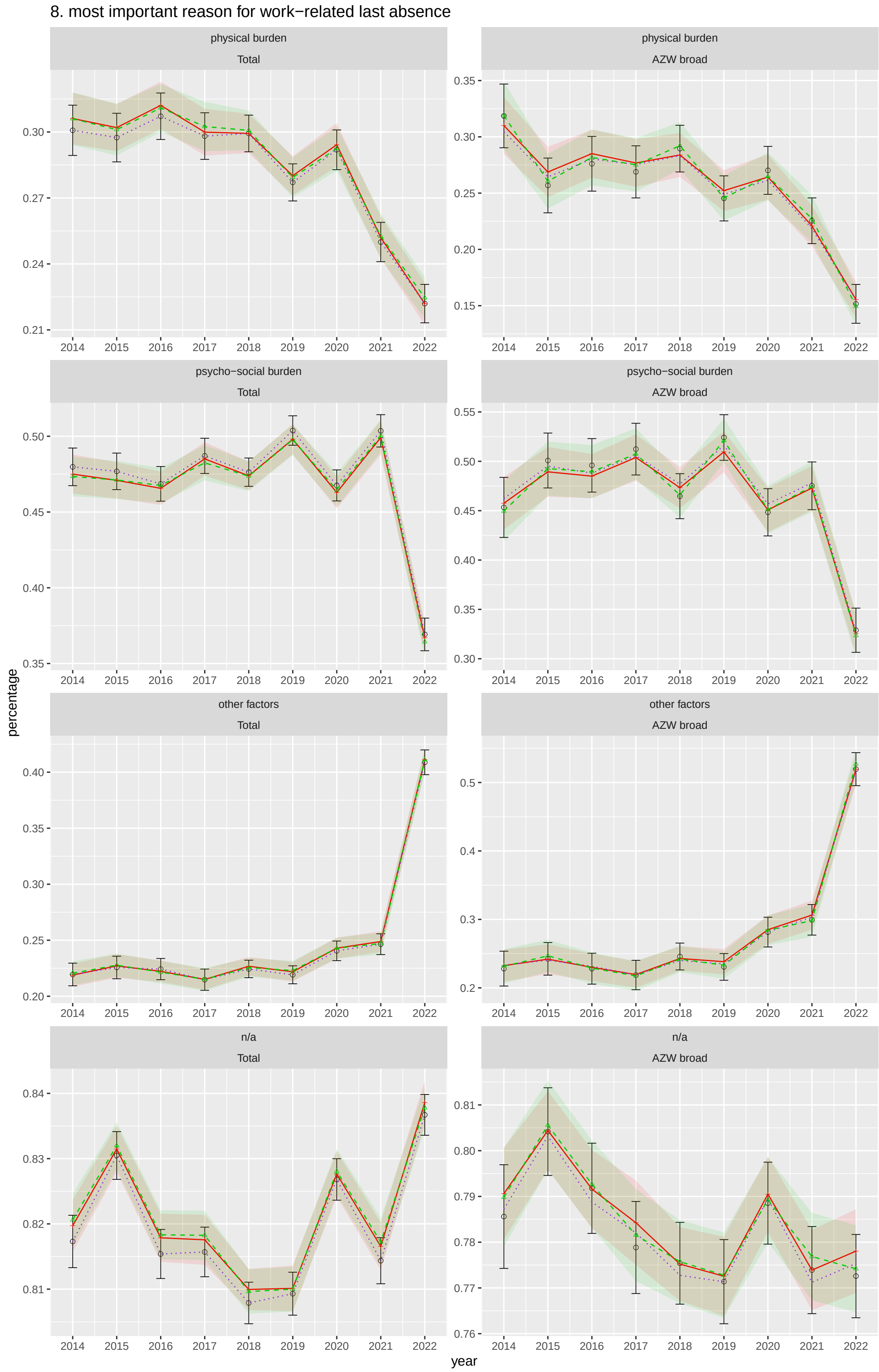

Figure 4.25 Estimates and approximate 95% confidence intervals/bands, for all employees (left) and for the AZW population (right) for all categories. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

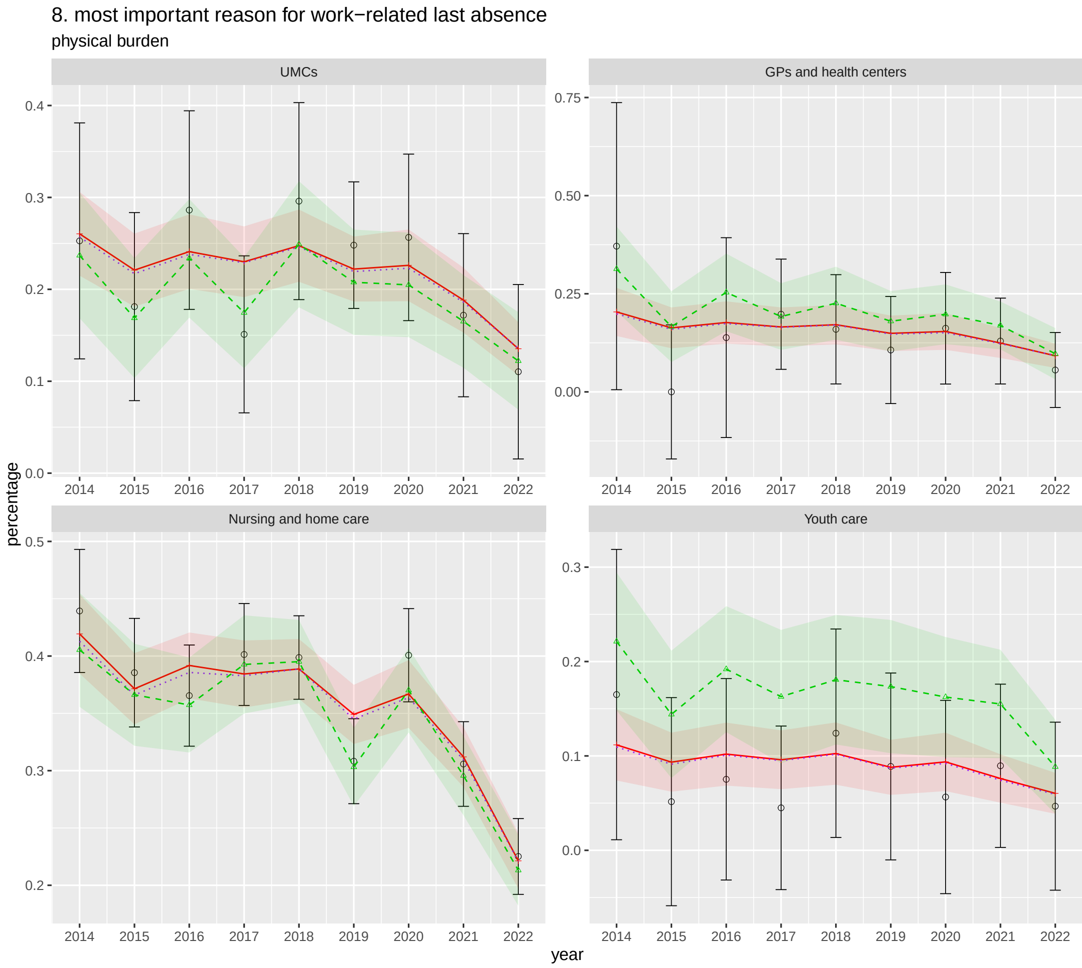

Figure 4.26 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 AZW branches. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

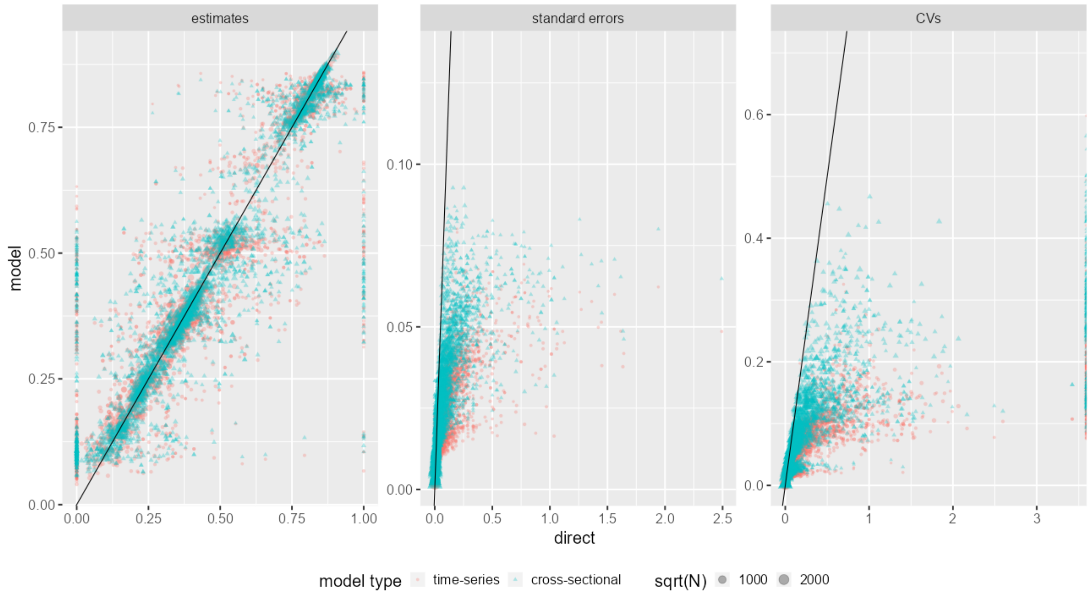

Figure 4.27 Time-series and cross-sectional model estimates versus direct estimates, standard errors and CVs for the complete set of 718 domains for the reason of work-related absence in 2022.

5 Discussion

A unit-level modelling approach is used to estimate eight sickness absence indicators for a large set of subpopulations of employees, where the focus lies on subpopulations in the Health and Well-Being sector, referred to as the AZW sector in this report. The most detailed domains of interest are at the level of AZW branches by RegioPlus region, where AZW branch is a subdivision of the AZW sector in 10 branches plus the non-AZW sector, and RegioPlus is a subdivision of The Netherlands into 28 sub-provincial regions. The number of observations in the annual NEA survey, which measures the required sickness absence, can be very small for many of these detailed domains. Direct estimates based on the survey weights are plagued by very large variances at this level of detail, and are even occasionally undefined due to zero respondents in small domains in certain years. We therefore also consider time-series modelling of the underlying absence target variables, using NEA data from 2014-2022, to allow borrowing strength over time and other dimensions.

Estimates based on time-series multilevel models are compared not only to the survey-weighted direct estimates, but also to estimates based on cross-sectional multilevel models. The latter are fitted separately to each year's NEA data. Although such models are not able to borrow strength over time, they can still exploit similarities between the various domains, such as regions and AZW branches. The results presented show that both time-series and cross-sectional model estimates are on average much more plausible and accurate than the direct estimates at the detailed levels of interest.

A careful model selection procedure has led to a time-series multilevel model that includes the most important fixed and random effects. The fixed effects include those variables used in the sampling design and weighting of the NEA survey. The random effects are selected to allow for variation over the most important dimensions of interest, corresponding to AZW branch, RegioPlus, age class, business size class, time (year) and several of their interactions. A single fixed and random effects specification is ultimately used for all eight indicators, although different families of sampling distributions are used to fit each target variable. In particular, four of the eight target variables are categorical, and here multinomial (time-series) multilevel models are used with the same fixed and random effects specification for each category.

It turns out that the time-series model estimates improve on the cross-sectional model estimates in several respects. First and foremost, the time-series estimates on average over all years remain closer to the average of the direct estimates, which is desirable as the direct estimates are approximately unbiased by design and as one averages over more years their variances decrease to more acceptable levels, except perhaps for the smallest domains. Second, the time-series model-based estimates are generally more smooth in their dependence over time, and this usually results in more plausible series of estimates. Also, estimated DIC and WAIC model criteria show a preference for the time-series models, for all eight indicators. Finally, the estimated uncertainty intervals of the time-series model estimates are mostly narrower compared to their cross-sectional model-based counterparts.

Eventually, the estimates based on the time-series model have been selected for publication. However, these estimates have first been benchmarked to be in agreement with the already published direct estimate at the overall level, for each indicator. This benchmarking step resulted in only small adjustments of the model-based small area estimates, especially at the more detailed levels. These newly estimated figures replace and extend the hitherto published three-year moving average estimates for the first four indicators. For most indicators, there are only few time-series model estimates with relative standard errors (CVs) exceeding 25%. Such estimates have been suppressed in the published tables.

The questionnaire of NEA has been (partly) redesigned in 2022. Effects of these changes can be seen especially in the time-series plots for the binary absence indicator. Apparently the time-series modelling did not suffer too much from these changes. However, this may hamper analyses of developments of sickness absence over time. At present there is no easy way to disentangle the discontinuities into real developments and changes in measurement level due to the redesign. It would, however, be possible to start estimating the effects due to the redesign after more years of NEA data have been collected under the new design, using a level intervention approach as proposed in Boonstra et al. (2021). A time-series model including such effects might then be used to correct the series of estimates to either the new or old measurement levels.

Acknowledgement

We wish to thank Henk-Jan Dirven, John Michiels, Marjolein Peters and Rob Willems for their useful comments that helped improve this report.

References

Battese, G., R. Harter, and W. Fuller (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical*

Association 83(401), 28–36.

Boonstra, H., S. Das, and J. van den Brakel (2021). Small area estimation of sickness absence based on the netherlands working conditions survey. CBS Discussion Paper, January 28, 2021.

Boonstra, H. J. (2021). *mcmcsae: Markov Chain Monte Carlo Small Area Estimation*. R package version 0.6.0.

Boonstra, H. J., B. Buelens, and M. Smeets (2007). Estimation of municipal unemployment fractions - a simulation study comparing different small area estimators. BPA nr. DMK-DMH-2007-04-20-HBTA, Statistics Netherlands.

Boonstra, H. J., J. Van den Brakel, and S. Das (2021). Multilevel time series modeling of mobility trends. *Journal of the Royal Statistical Society Series A*, 985–1007.

CBS and TNO (2023). Nationale enquête arbeidsomstandigheden (nea) 2022 - onderzoeksbeschrijving. Methodological report, in Dutch, https://www.cbs.nl/nl-nl/longread/rapportages/2023/nationale-enquete-arbeidsomstandigheden--nea---2022-onderzoeksbeschrijving.

de Vries, J. and J. Michiels (2019). Haalbaarheidsonderzoek kleinedomeinschatters positie in de werkkring en onderwijsdeelname. report version 17-12-2019, in Dutch, Statistics Netherlands.

Gao, Y., L. Kennedy, D. Simpson, and A. Gelman (2021). Improving multilevel regression and poststratification with structured priors. *Bayesian Analysis 16*(3), 719.

Gelfand, A. and A. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515–533.

Gelman, A., D. V. Dyk, Z. Huang, and W. Boscardin (2008). Using redundant parameterizations to fit hierarchical models. *Journal of Computational and Graphical Statistics 17*(1), 95–122.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., J. Lax, J. Phillips, J. Gabry, and R. Trangucci (2016). Using multilevel regression and poststratification to estimate dynamic public opinion. *Unpublished manuscript, Columbia University 2*.

Gelman, A., X.-L. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 733–760.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*(4), 457–472.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell. 6*, 721–741.

Linderman, S., M. J. Johnson, and R. P. Adams (2015). Dependent multinomial models made easy: Stick-breaking with the pólya-gamma augmentation. *Advances in Neural Information Processing Systems 28*.

O'Malley, A. and A. Zaslavsky (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association 103*(484), 1405–1418.

Polson, N. G., J. G. Scott, and J. Windle (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association 108*(504), 1339–1349.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, J. and I. Molina (2015). *Small Area Estimation*. Wiley-Interscience.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B 64*(4), 583–639.

van Dam, L., G. Mars, J. Knops, S. van den Heuvel, E. de Vroome, A. Pleijers, M. Ramaekers, and B. Janssen (2022). Nationale enquête arbeidsomstandigheden 2021. Methodological report, in Dutch, TNO and CBS.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing 27*, 1413–1432.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research 11*, 3571–3594.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research 14*, 867–897.

Zhou, M. and L. Carin (2015). Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*(2), 307–320.

Zhou, M., L. Li, D. Dunson, and L. Carin (2012). Lognormal and gamma mixed negative binomial regression. In *Machine learning: proceedings of the International Conference. International Conference on Machine Learning*, Volume 2012, pp. 1343. NIH Public Access.

Appendix

A Covariates used in the multilevel models

Variable	Description	Categories
sex	Gender	male , female
ageclass3	Age in 3 classes	15-34, 35-54, 55-74
ageclass6	Age in 6 classes	15-24, 25-34, 35-44, 45-54, 55-64, 65-74
background	Migrant background	native, 1st generation western, 2nd generation western, 1st generation nonwestern 2nd generation nonwestern
nonwestern	background in 2 classes	nonwestern, other
Stratum	Sampling design strata	39 strata
Oversampled	Oversampled educational branches	0,1,41-46
Region	The Netherlands in 4 regions	North, East, West, South
Province	Province	12 provinces
RegioPlus	RegioPlus labour market region	28 regions
Urban	Degree of urbanisation	extremely urbanised, strongly urbanised moderately urbanised, slightly urbanised not urbanised
AZW	Health and non-health dichotomy	AZW (broad), non-AZW
AZWbranch	Health sector subdivided into 10 classes	see Appendix B
sizeclass9	Employer's size class in 9 groups	size class 1 - 9
sizeclass3	Employer's size class in 3 groups	small, medium, large
hhtype	Household type	single household, unmarried couple without children + unknown, married couple without children, unmarried couple with children, married couple with children, single parent household
income	Income in 6 decile groups	1st decile + unknown, 2nd decile decile 3 and 4, decile 5 and 6 decile 7 and 8, decile 9 and 10
contract	Type of contract	permanent, temporary
Year	year in 9 classes	2014-2022

B AZW branch names in English and Dutch

1	UMCs	Universitair Medische Centra
2	Hospitals and other MCs	Ziekenhuizen en overige medisch specialistische zorg
3	Mental Health Care	Geestelijke gezondheidszorg
4	GPs and health centers	Huisartsen en gezondheidscentra
5	Other care	Overige zorg
6	Nursing and home care	Verpleging, verzorging en thuiszorg
7	Disability care	Gehandicaptenzorg
8	Childcare	Kinderopvang (inclusief peuterspeelzaalwerk)
9	Youth care	Jeugdzorg
10	Social care	Sociaal werk
11	Non-AZW	Niet-AZW

C Additional results

C.1 Percentage absence time

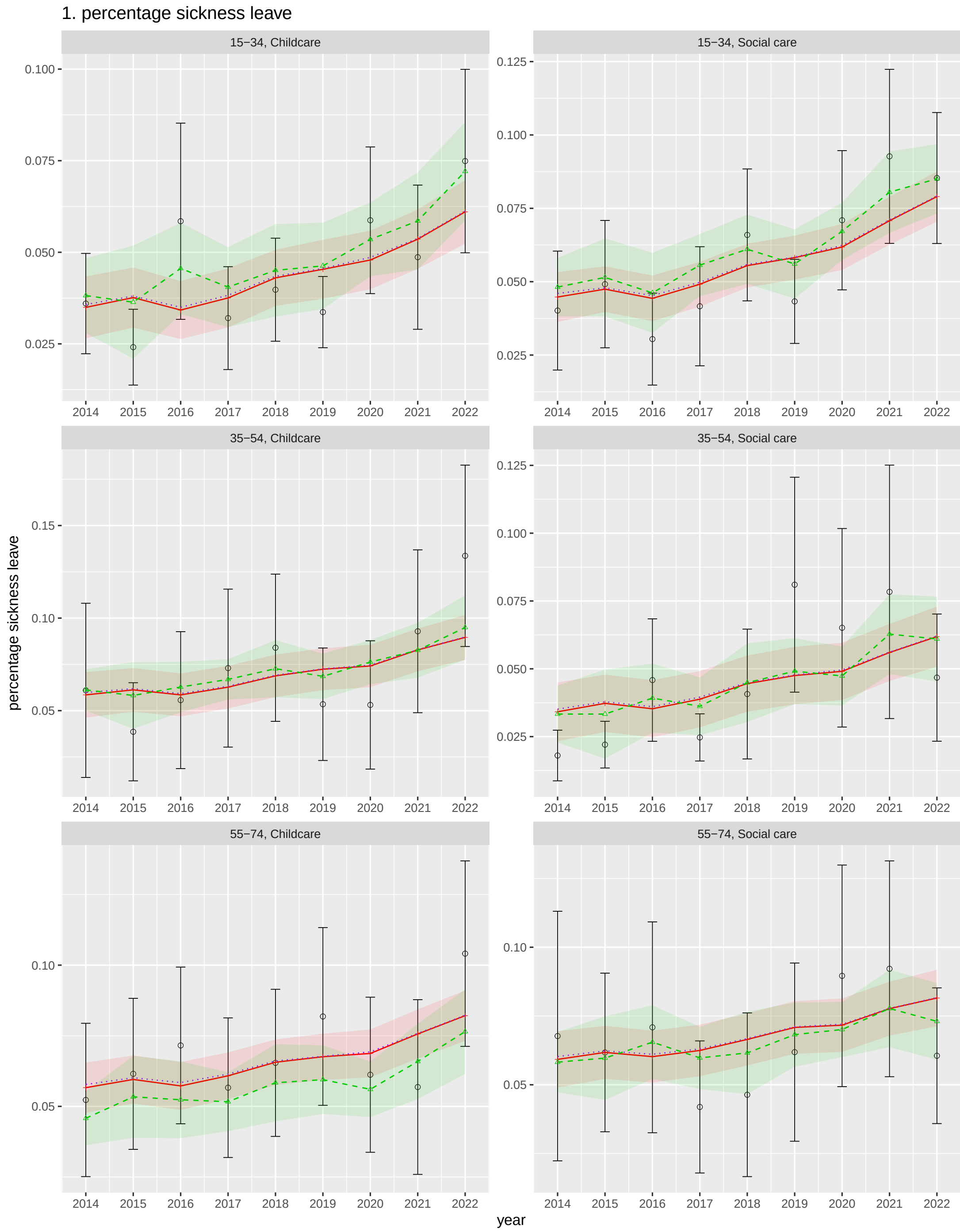

Figure C.1 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

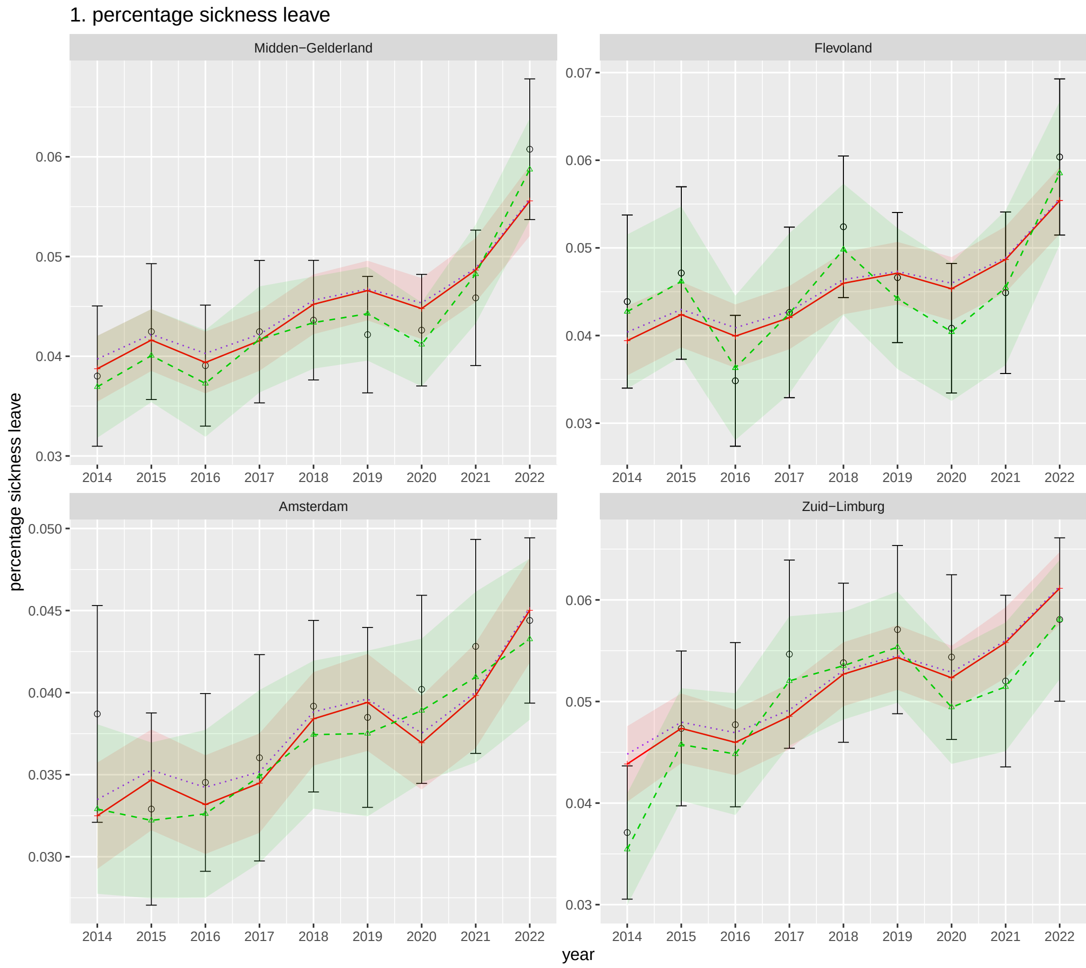

Figure C.2 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

C.2 Percentage of employees with absence

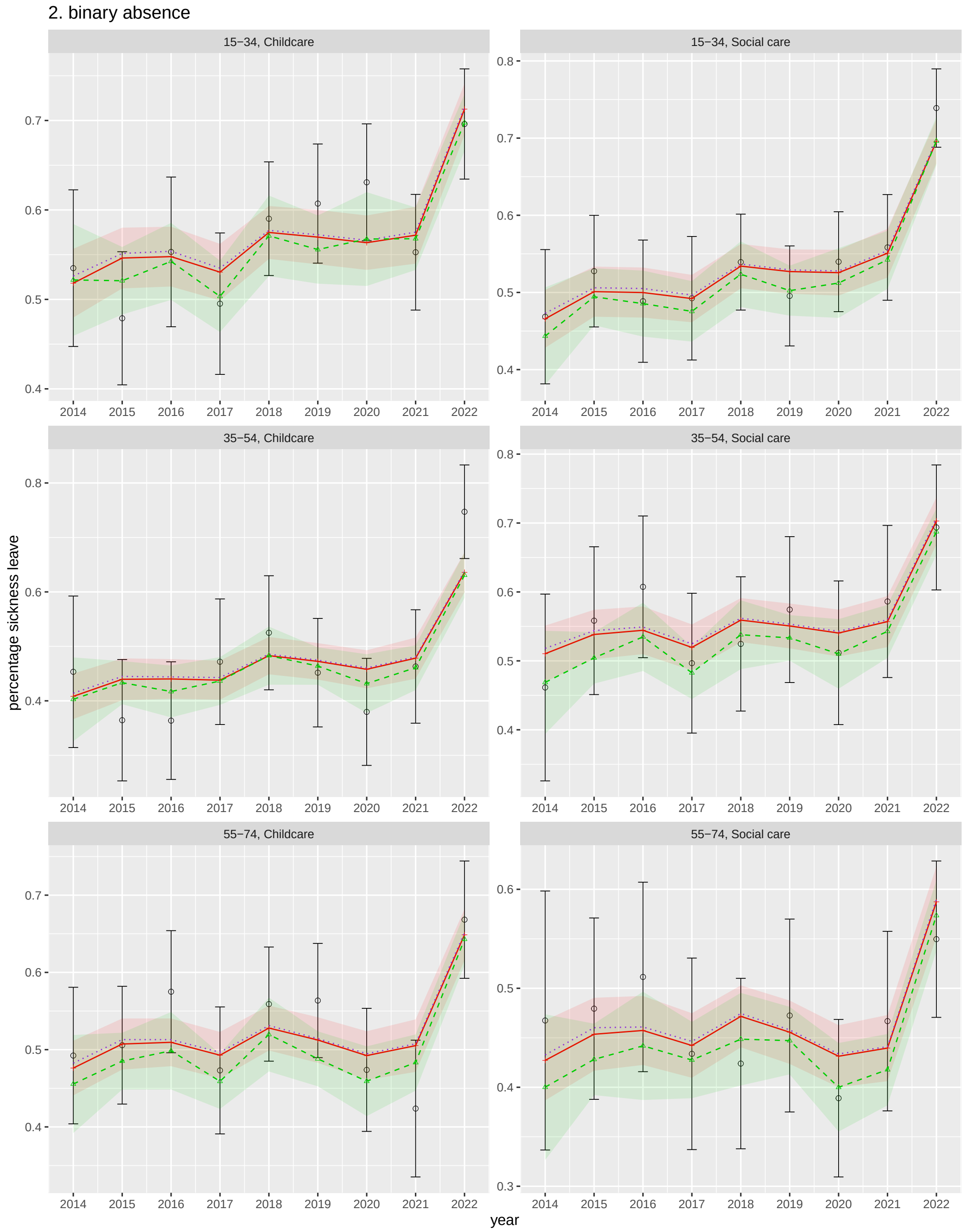

Figure C.3 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

2. binary absence

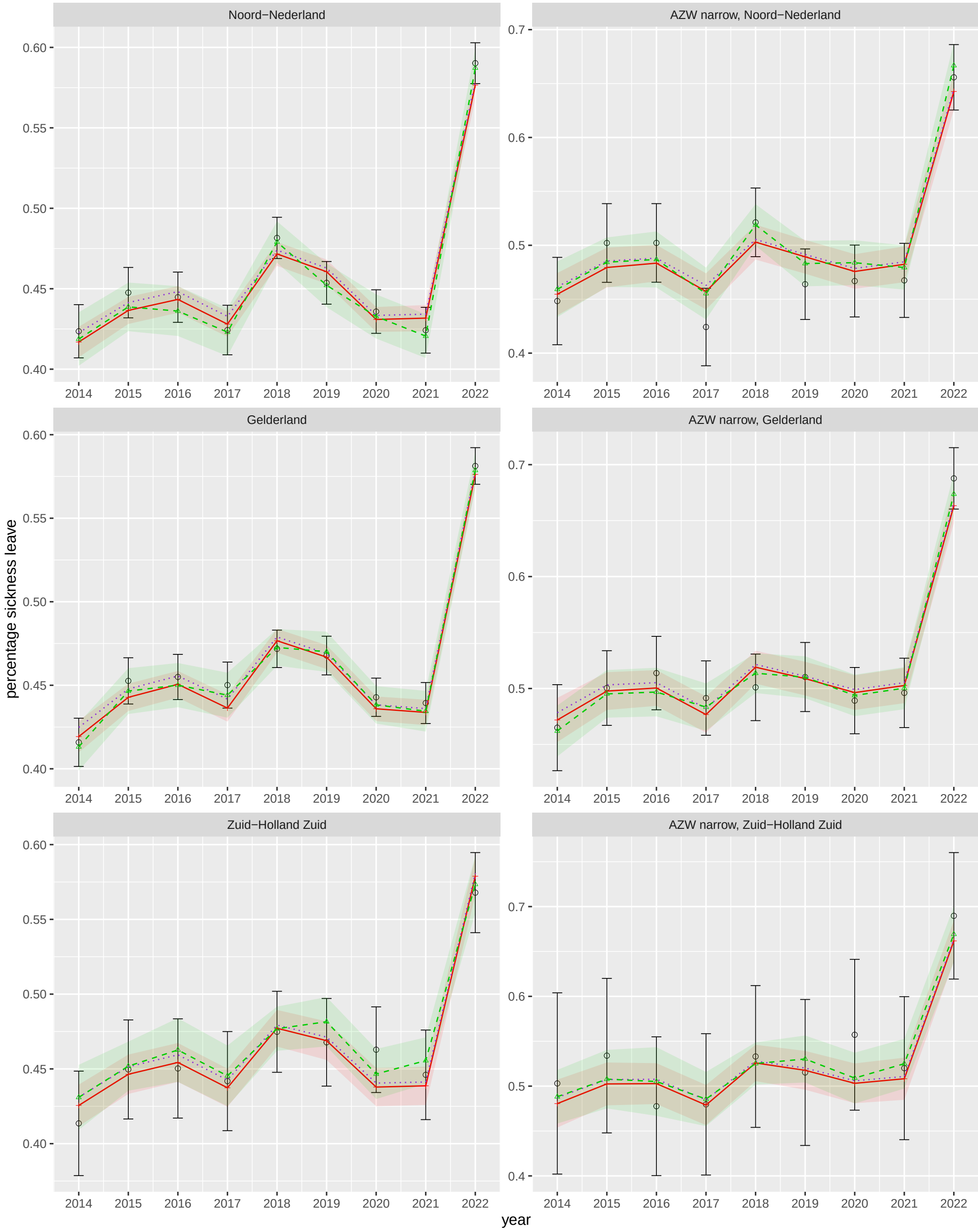

Figure C.4 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

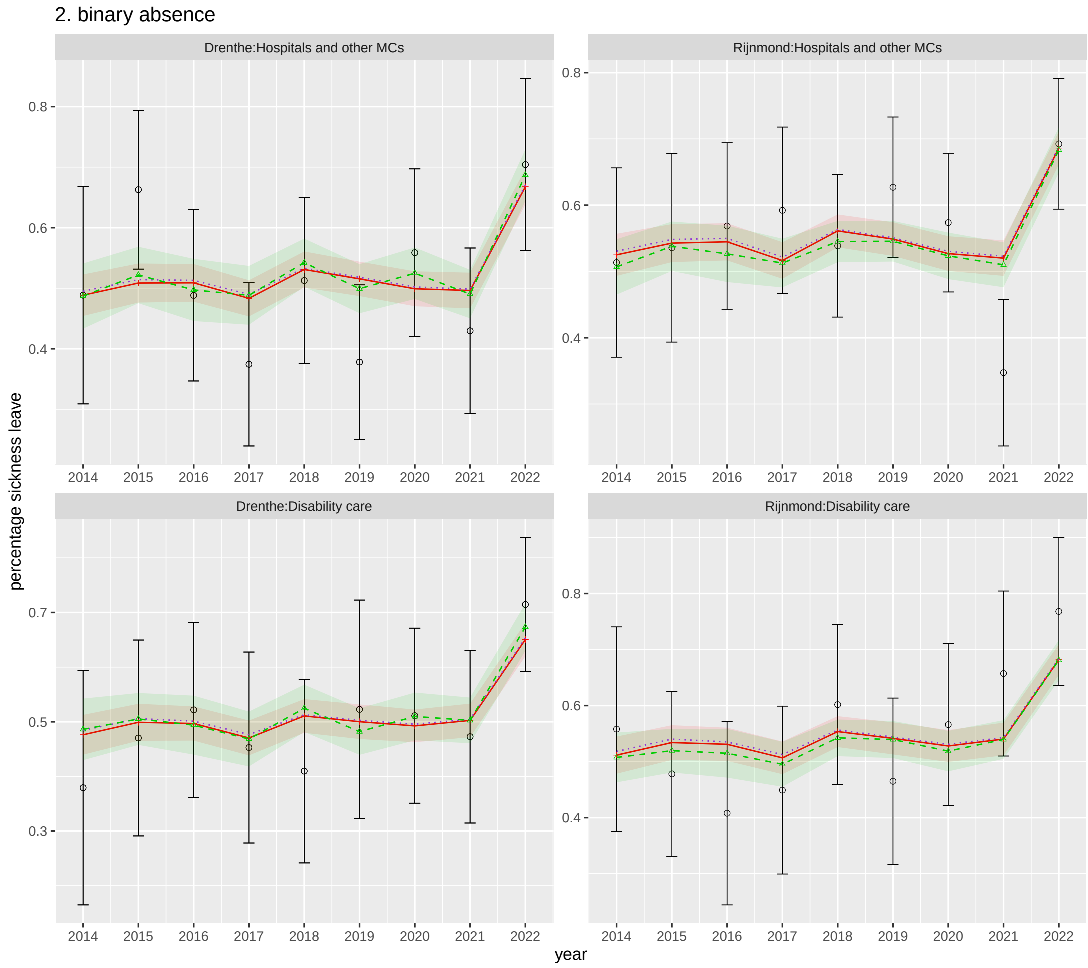

Figure C.5 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-4.5	-0.8	-0.5	-0.6	-0.2	2.0
	cross-sectional	-4.2	-1.1	-0.6	-0.6	-0.2	2.0
AZW branch	time-series	-6.6	-1.3	-0.7	-0.5	0.1	5.2
	cross-sectional	-9.5	-3.6	-1.1	-1.0	1.0	10.8
region	time-series	-1.5	-0.9	-0.6	-0.5	-0.4	2.6
	cross-sectional	-2.2	-1.0	-0.8	-0.5	-0.5	3.8
RegioPlus	time-series	-3.6	-1.5	-0.8	-0.6	0.2	2.6
	cross-sectional	-3.3	-2.1	-0.7	-0.6	0.3	3.9
region-branch	time-series	-20.9	-2.5	-0.5	0.2	2.3	49.6
	cross-sectional	-21.3	-3.9	-0.6	-0.2	2.4	42.8
RegioPlus-branch	time-series	-45.3	-4.8	-0.4	1.0	4.7	181.7
	cross-sectional	-44.2	-5.4	-0.5	0.6	4.2	181.6

Table C.1 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-161.0	-50.2	-30.7	-34.9	-9.9	73.6
	cross-sectional	-168.6	-56.8	-36.8	-38.5	-6.1	74.1
AZW branch	time-series	-155.2	-21.6	-11.5	-16.4	0.3	68.3
	cross-sectional	-170.1	-53.3	-23.2	-24.6	21.4	78.2
region	time-series	-127.5	-31.1	-21.6	-25.6	-11.4	41.9
	cross-sectional	-128.7	-49.7	-28.3	-28.4	-15.6	60.7
RegioPlus	time-series	-61.5	-36.2	-17.5	-15.6	2.6	41.9
	cross-sectional	-111.7	-41.3	-18.7	-16.6	5.3	102.7
region-branch	time-series	-125.5	-28.8	-5.0	-7.2	16.4	83.5
	cross-sectional	-124.6	-38.6	-8.3	-11.5	14.9	76.6
RegioPlus-branch	time-series	-104.6	-27.2	-3.4	-4.6	17.5	83.5
	cross-sectional	-115.9	-30.6	-4.4	-6.8	17.2	118.5

Table C.2 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-2.7	18.8	26.7	28.1	39.8	59.7
	cross-sectional	-4.2	0.4	2.0	8.0	11.2	31.1
AZW branch	time-series	-1.8	46.5	58.4	53.6	66.5	75.4
	cross-sectional	-4.9	31.7	46.4	42.5	56.4	70.6
region	time-series	13.6	38.7	50.5	48.3	57.9	73.3
	cross-sectional	-0.9	2.0	33.4	25.1	39.2	52.6
RegioPlus	time-series	43.3	55.3	62.2	61.0	65.9	77.3
	cross-sectional	1.2	30.8	47.2	42.1	52.3	70.2
region-branch	time-series	11.4	69.9	77.5	73.7	84.1	92.5
	cross-sectional	1.9	61.5	70.8	65.5	79.3	91.2
RegioPlus-branch	time-series	42.3	79.2	84.2	81.1	87.5	92.5
	cross-sectional	7.3	73.1	79.7	74.7	85.0	91.2

Table C.3 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

C.3 Number of absent periods

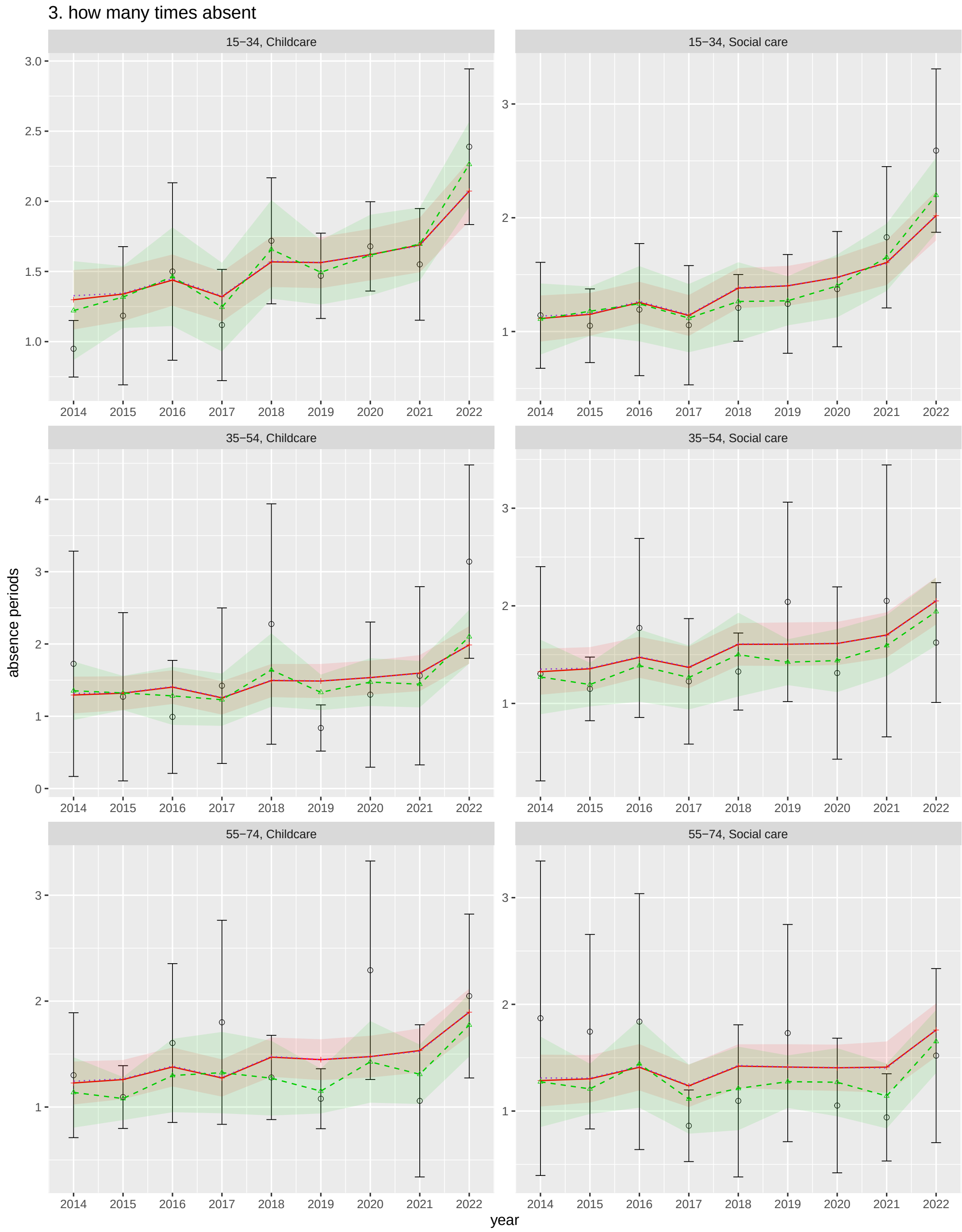

Figure C.6 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

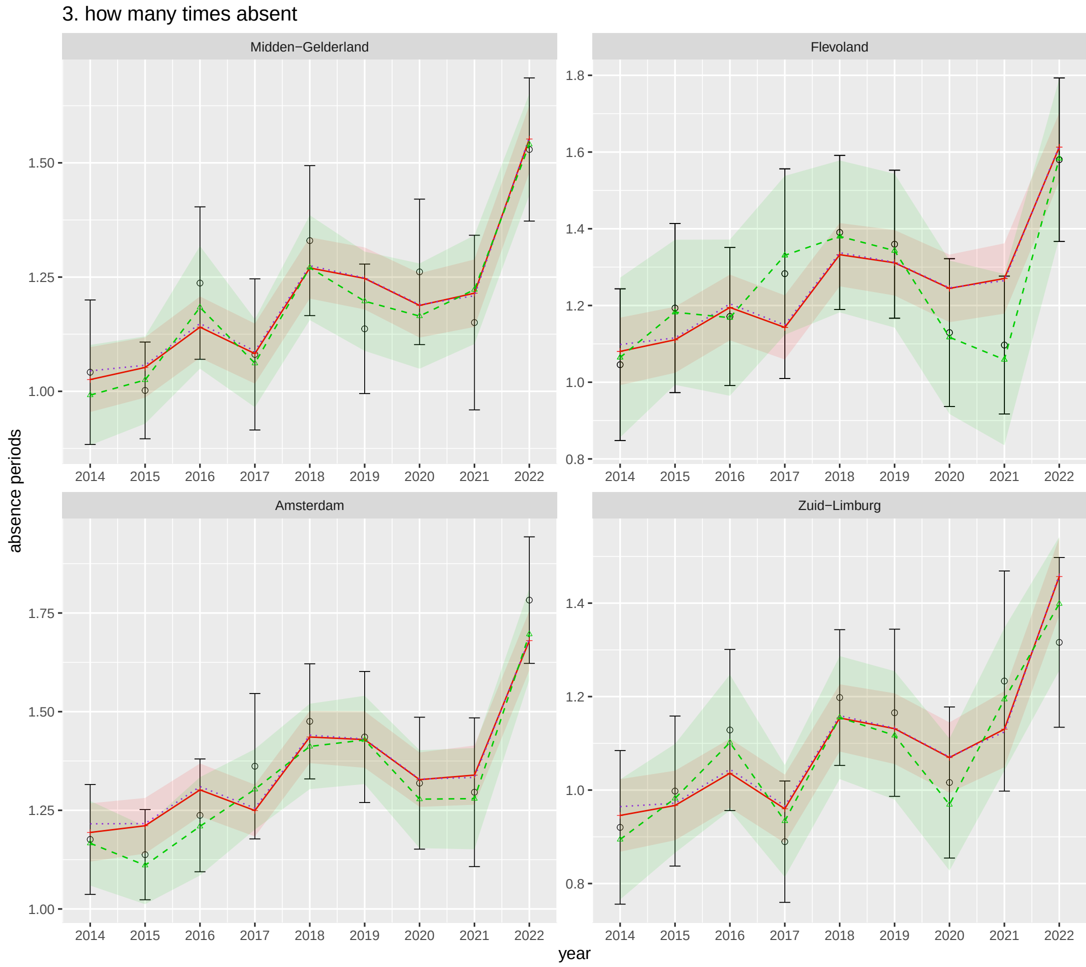

Figure C.7 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-5.5	-0.6	-0.2	-0.3	0.4	3.2
	cross-sectional	-4.4	-0.7	-0.3	-0.3	0.4	3.4
AZW branch	time-series	-12.4	-2.8	-0.2	-0.1	1.6	39.1
	cross-sectional	-22.0	-7.5	-0.5	-1.3	3.3	31.7
region	time-series	-3.7	-0.7	-0.2	0.3	0.5	11.1
	cross-sectional	-4.8	-1.2	-0.4	0.3	1.8	9.9
RegioPlus	time-series	-11.8	-2.0	-0.5	-0.2	1.5	11.1
	cross-sectional	-10.9	-2.2	-0.2	0.0	1.9	10.8
region-branch	time-series	-60.1	-5.8	-0.1	4.8	8.6	147.0
	cross-sectional	-64.8	-9.1	-0.5	3.8	8.7	165.5
RegioPlus-branch	time-series	-61.2	-7.8	0.5	8.9	19.0	327.8
	cross-sectional	-65.3	-10.8	0.9	7.8	18.5	295.3

Table C.4 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

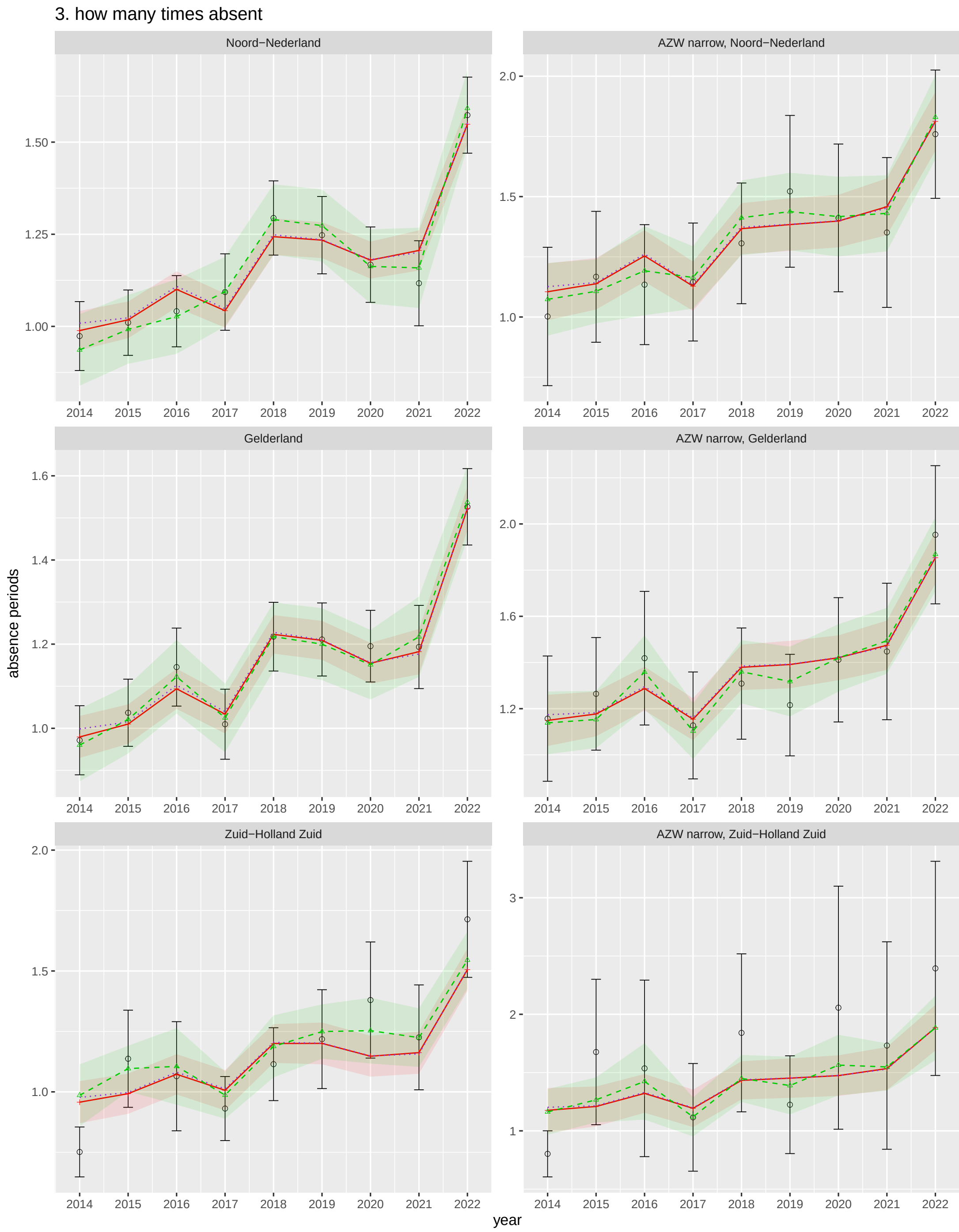

Figure C.8 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

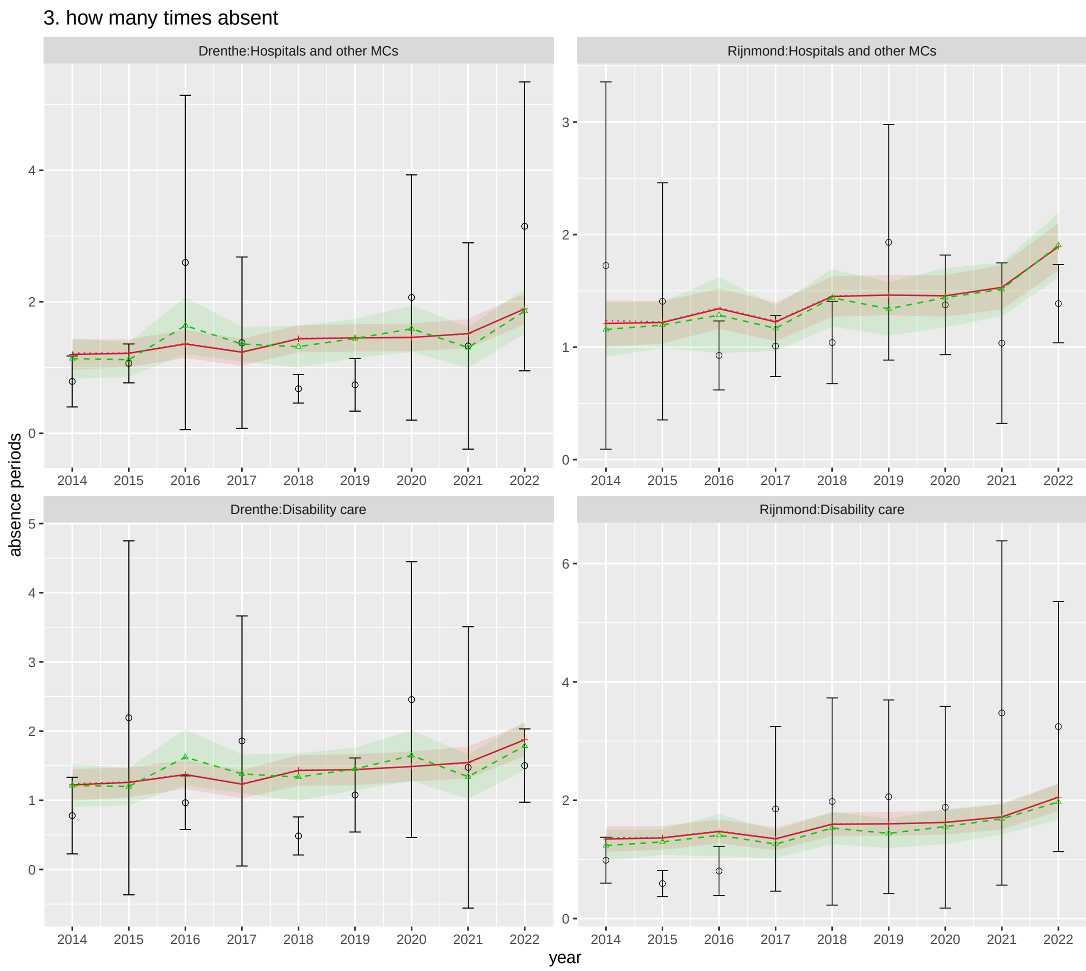

Figure C.9 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-57.2	-15.8	-4.5	-3.1	10.9	49.7
	cross-sectional	-46.2	-23.9	-8.3	-4.1	11.7	52.4
AZW branch	time-series	-37.9	-13.8	-2.1	-1.5	9.5	76.7
	cross-sectional	-68.9	-32.8	-8.4	-6.0	18.5	69.7
region	time-series	-49.1	-10.7	-2.6	-2.9	8.8	48.7
	cross-sectional	-50.2	-19.8	-5.8	-3.0	12.3	54.8
RegioPlus	time-series	-53.0	-15.1	-4.5	-2.7	11.4	48.7
	cross-sectional	-49.0	-22.4	-1.6	-1.9	11.8	49.8
region-branch	time-series	-73.1	-16.1	-2.0	8.3	26.4	222.3
	cross-sectional	-75.3	-21.1	-3.7	6.3	28.1	186.5
RegioPlus-branch	time-series	-73.1	-16.6	0.9	12.7	30.7	222.3
	cross-sectional	-77.1	-21.9	2.6	11.4	33.2	254.7

Table C.5 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-2.7	25.3	35.5	33.7	46.9	53.8
	cross-sectional	-18.6	-6.2	11.8	7.4	16.1	33.3
AZW branch	time-series	-3.7	56.8	63.5	60.0	71.1	80.1
	cross-sectional	-21.5	34.3	46.1	42.8	58.2	71.9
region	time-series	18.8	49.3	58.3	54.7	64.0	73.0
	cross-sectional	-13.2	1.8	37.4	27.1	43.6	54.0
RegioPlus	time-series	42.4	56.7	64.6	63.6	69.1	80.1
	cross-sectional	-13.2	32.1	47.8	42.3	54.2	73.3
region-branch	time-series	16.3	67.0	75.2	72.2	79.9	95.1
	cross-sectional	-7.3	54.2	65.8	59.8	72.5	94.1
RegioPlus-branch	time-series	33.2	66.9	75.6	73.8	82.0	95.1
	cross-sectional	-7.1	56.8	67.9	64.3	76.1	93.6

Table C.6 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

C.4 Number of absent days

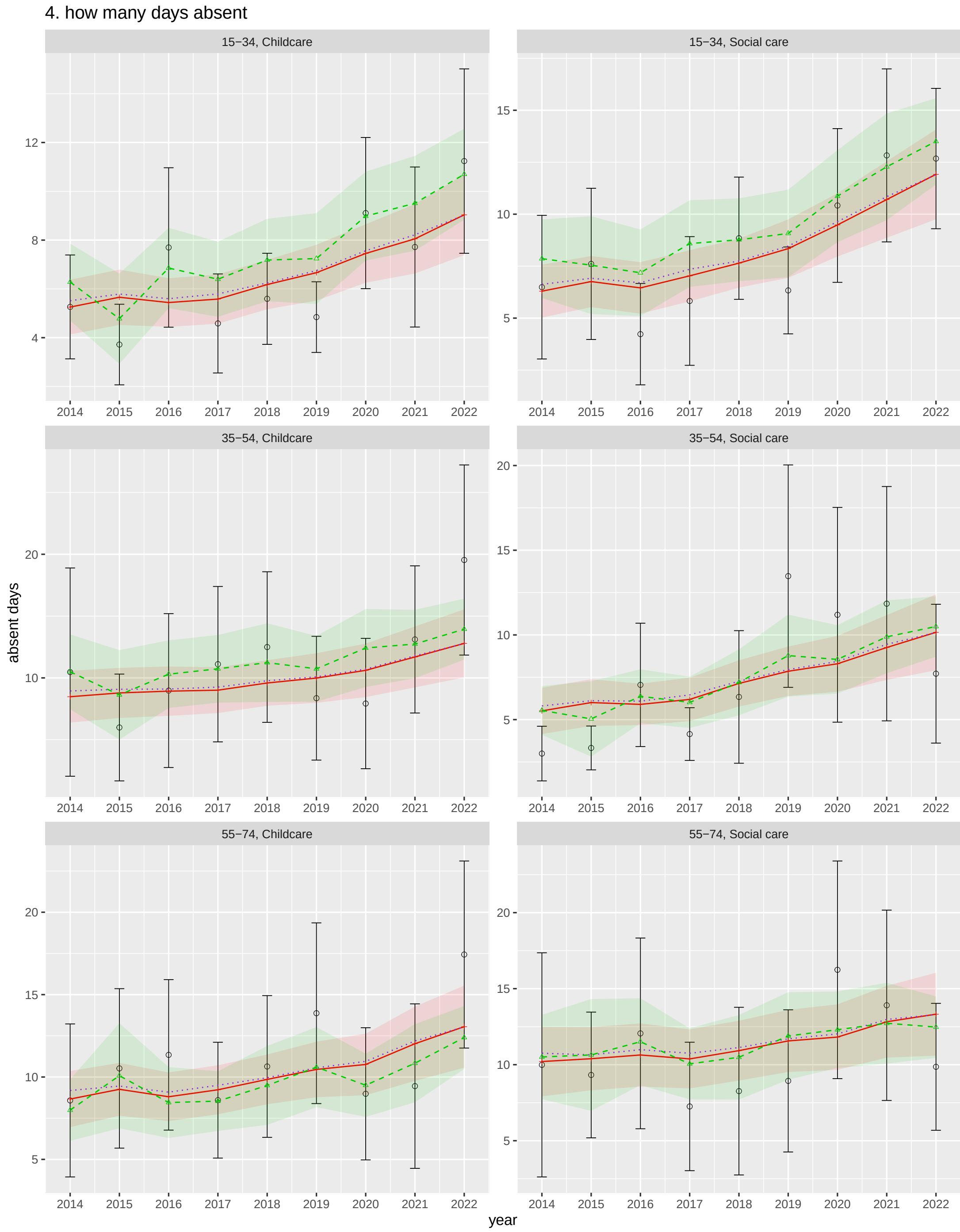

Figure C.10 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

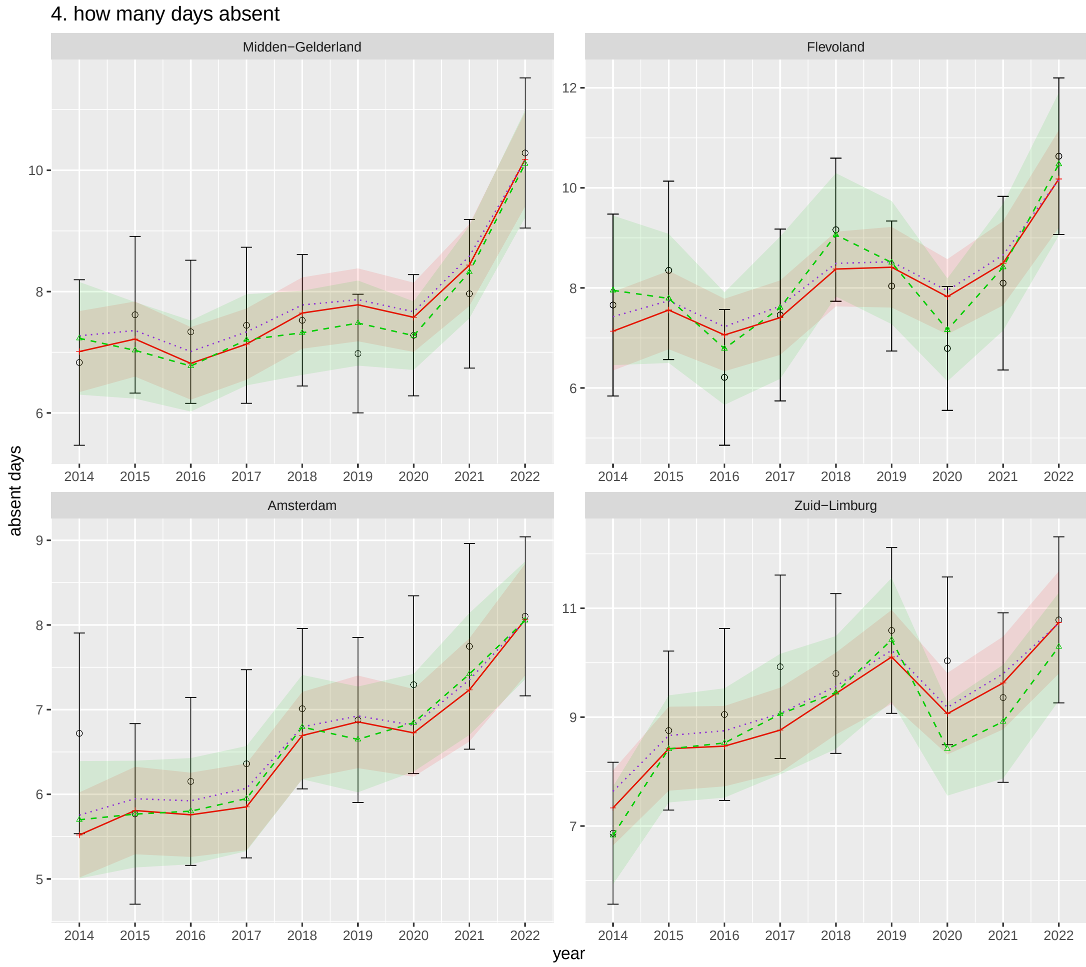

Figure C.11 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-12.3	-2.3	-1.0	-1.7	-0.2	6.3
	cross-sectional	-11.5	-1.5	0.0	-0.6	0.8	7.9
AZW branch	time-series	-19.5	-2.8	-1.1	-0.1	2.7	15.5
	cross-sectional	-23.3	-4.5	0.6	0.6	6.1	18.5
region	time-series	-5.6	-2.6	-1.5	-1.1	0.1	5.9
	cross-sectional	-8.5	-2.2	-0.6	0.4	2.7	8.6
RegioPlus	time-series	-9.9	-3.0	-1.2	-1.0	1.0	6.6
	cross-sectional	-17.4	-2.3	0.3	0.9	4.9	20.8
region-branch	time-series	-51.4	-8.0	0.1	6.2	8.6	329.7
	cross-sectional	-46.4	-7.8	0.9	7.1	12.9	317.4
RegioPlus-branch	time-series	-67.9	-9.9	0.2	10.6	16.5	329.7
	cross-sectional	-72.0	-10.3	1.0	12.6	22.0	317.4

Table C.7 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

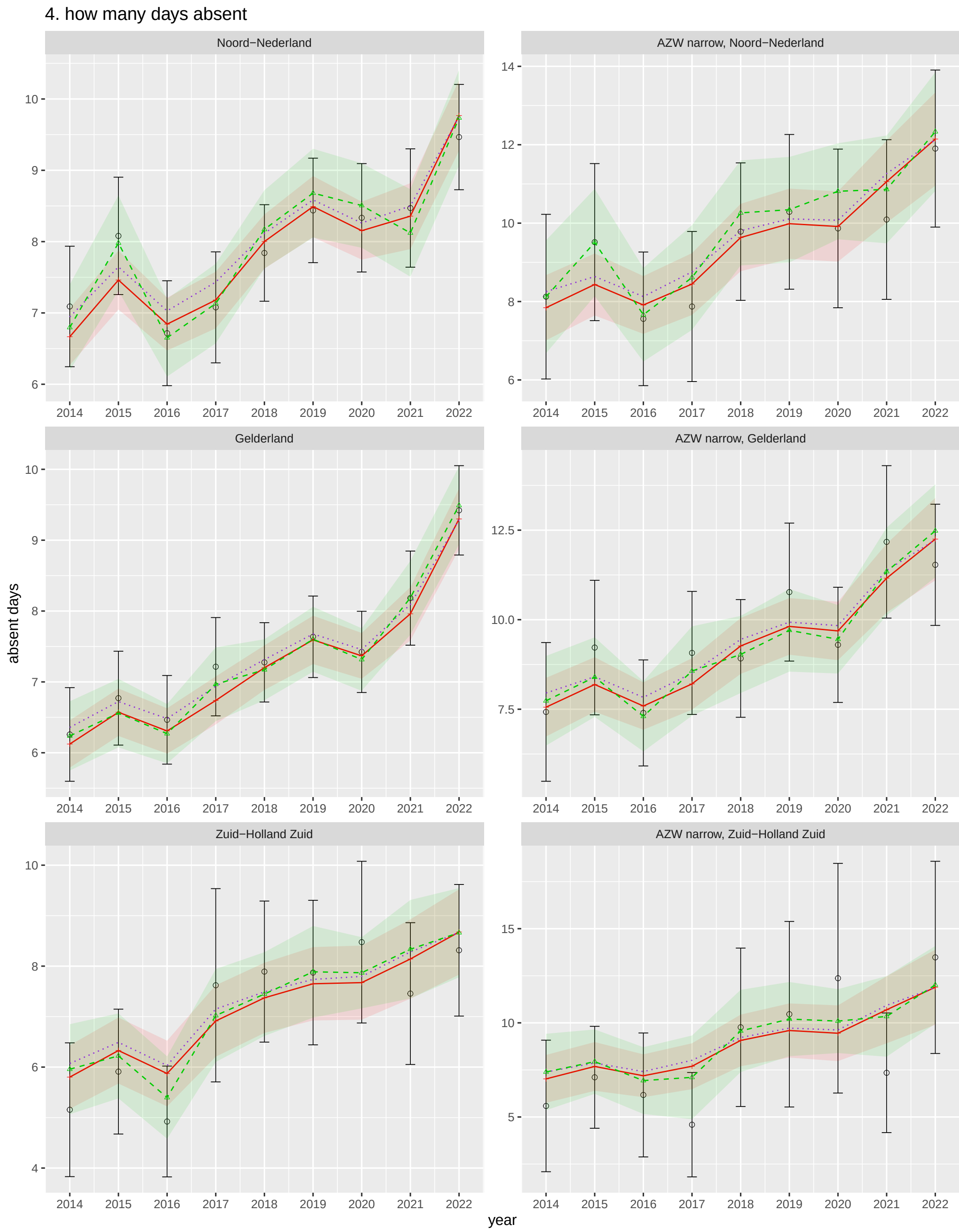

Figure C.12 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

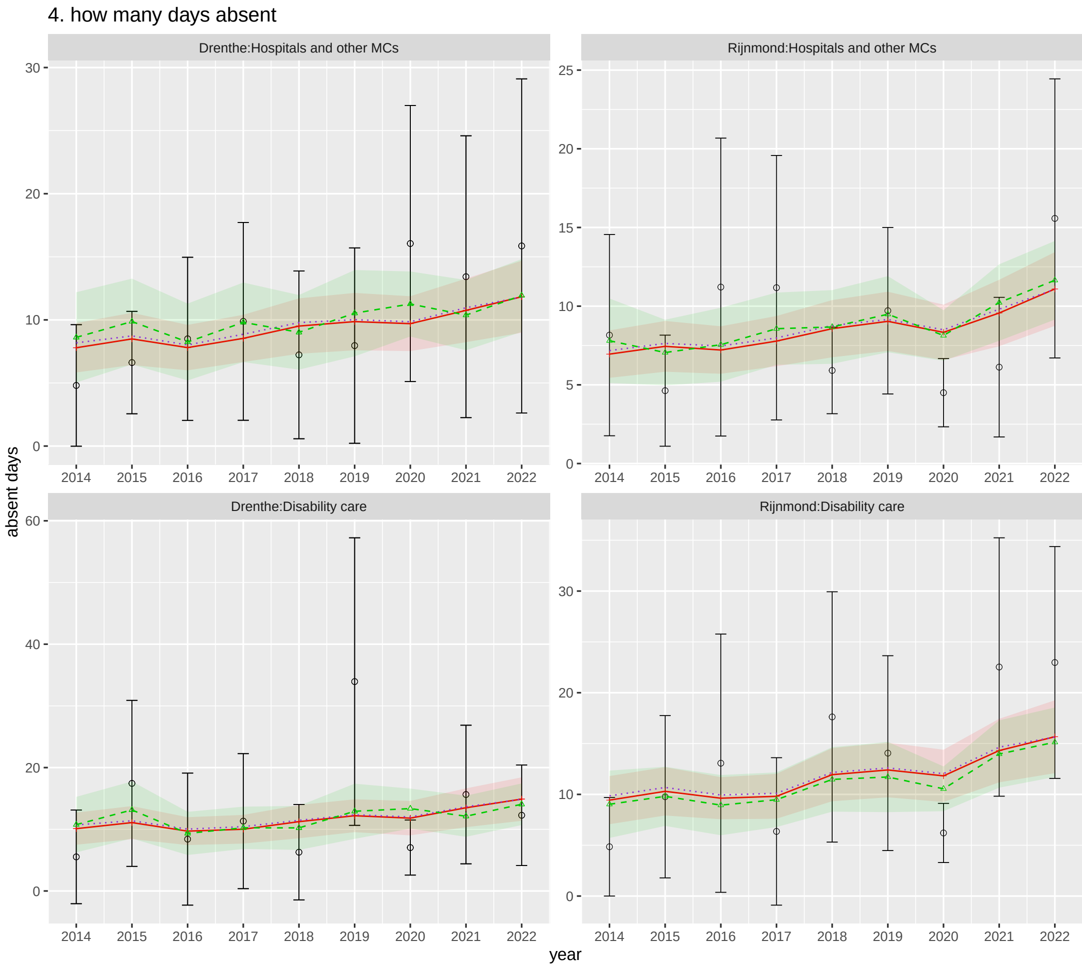

Figure C.13 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-120.3	-46.5	-29.8	-33.7	-3.0	72.5
	cross-sectional	-90.8	-30.3	0.7	-9.2	14.2	91.2
AZW branch	time-series	-120.1	-22.8	-4.9	-9.7	15.9	66.0
	cross-sectional	-101.6	-23.3	3.0	0.5	25.2	95.0
region	time-series	-99.0	-35.6	-16.4	-18.6	1.7	24.5
	cross-sectional	-81.7	-26.3	-10.9	-2.5	25.2	76.8
RegioPlus	time-series	-76.1	-26.0	-8.0	-10.5	7.9	26.5
	cross-sectional	-108.0	-15.9	2.5	0.3	26.0	59.3
region-branch	time-series	-98.5	-24.2	-0.3	0.4	20.4	144.4
	cross-sectional	-87.2	-23.6	2.5	5.9	32.7	180.0
RegioPlus-branch	time-series	-89.9	-22.6	-2.6	7.7	22.0	437.9
	cross-sectional	-89.3	-21.7	1.9	12.4	34.2	487.7

Table C.8 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	15.7	26.2	34.4	38.7	51.4	74.4
	cross-sectional	3.6	12.1	18.9	21.9	28.9	57.2
AZW branch	time-series	19.9	48.8	59.2	56.0	65.7	77.9
	cross-sectional	14.8	31.2	47.8	45.8	58.4	74.2
region	time-series	31.6	44.3	49.6	50.6	57.4	66.9
	cross-sectional	15.4	23.6	29.6	28.4	32.8	40.9
RegioPlus	time-series	41.0	49.8	60.0	57.1	61.3	73.0
	cross-sectional	17.3	33.4	39.2	39.0	44.7	58.1
region-branch	time-series	32.6	63.0	70.9	68.9	77.0	92.2
	cross-sectional	19.5	51.7	62.7	61.1	72.3	90.6
RegioPlus-branch	time-series	-26.0	65.5	72.0	69.5	78.5	97.7
	cross-sectional	-38.6	56.2	65.2	62.8	74.7	97.4

Table C.9 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

C.5 Duration of the last absence

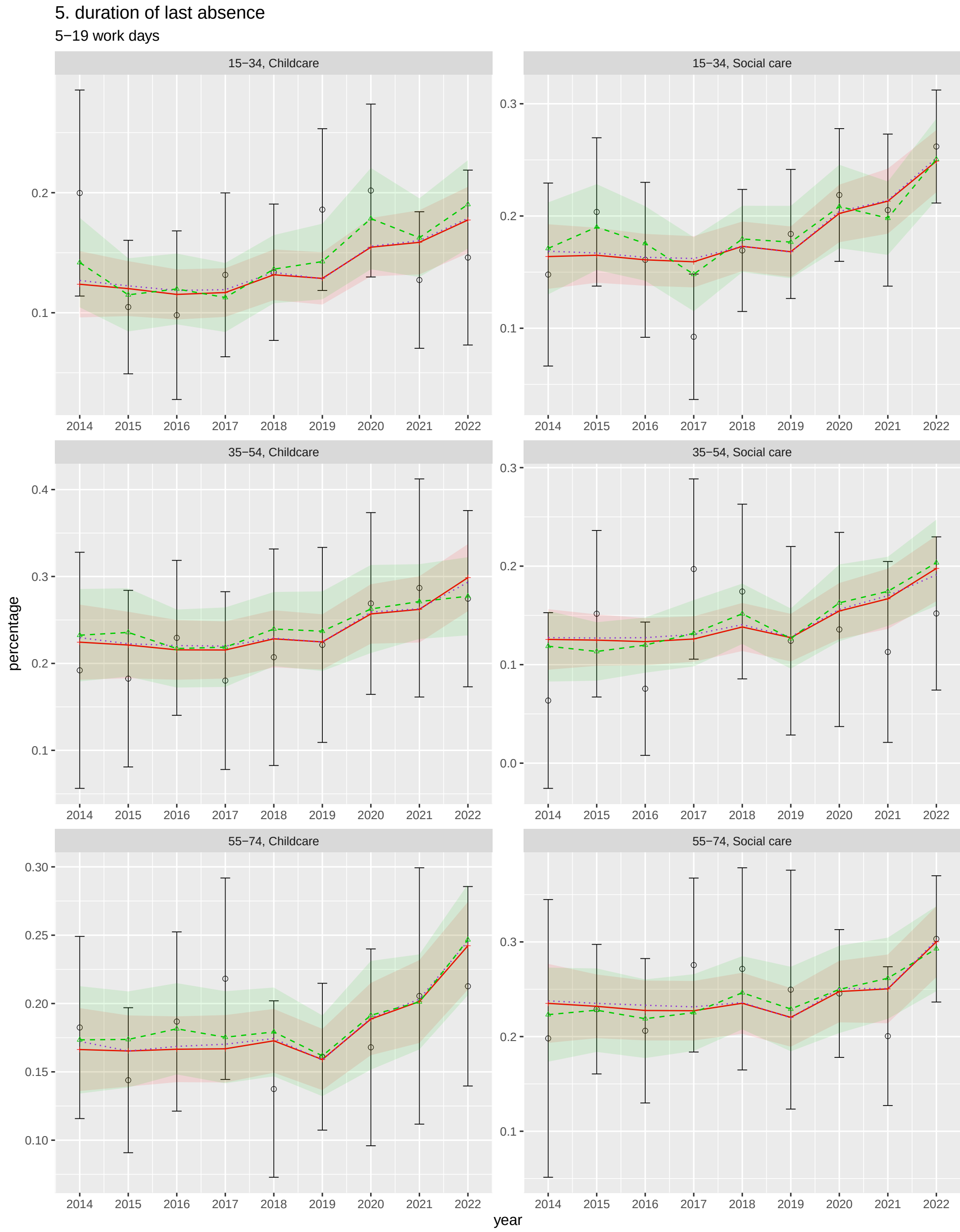

Figure C.14 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

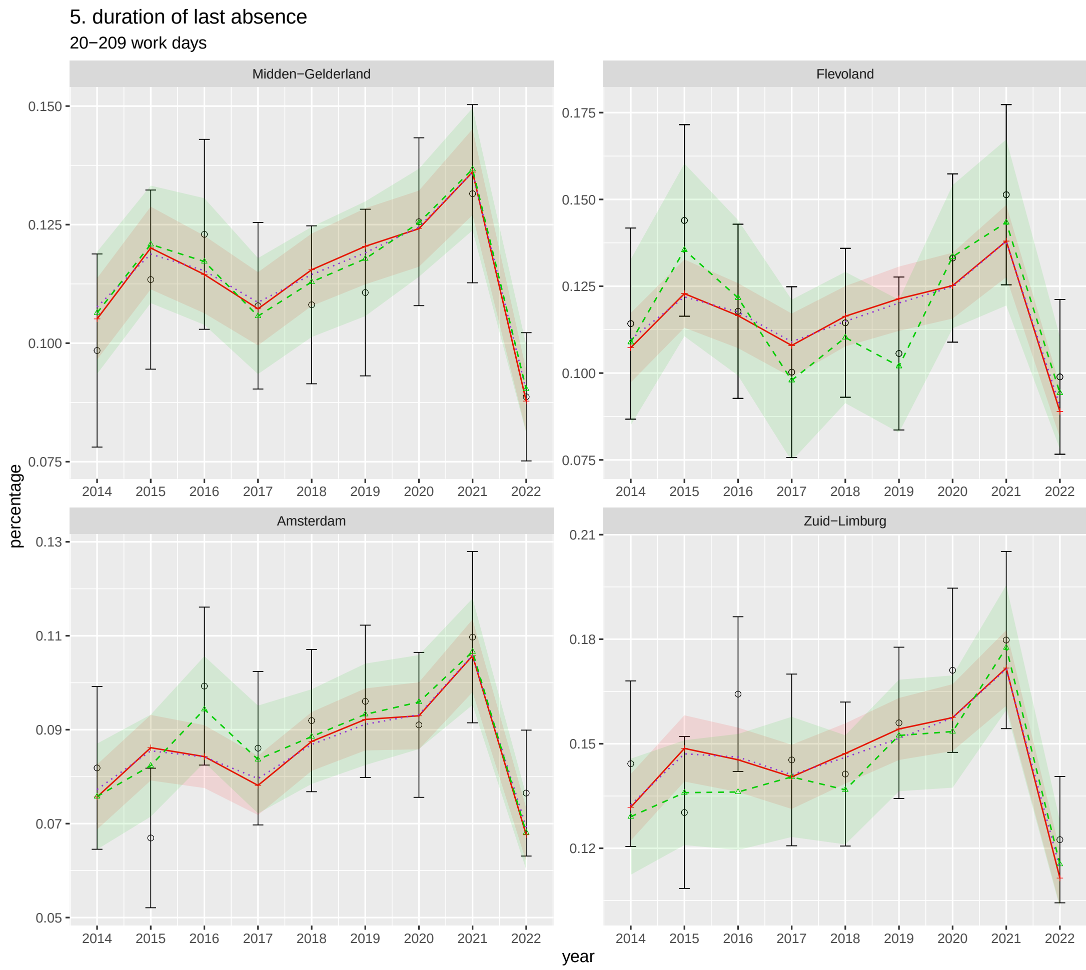

Figure C.15 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-12.0	-1.1	-0.0	-0.1	0.9	8.8
	cross-sectional	-13.0	-1.1	-0.0	-0.2	0.9	5.7
AZW branch	time-series	-29.7	-2.1	0.3	2.6	4.2	177.0
	cross-sectional	-39.7	-3.0	0.5	4.6	5.1	231.4
region	time-series	-14.9	-1.0	0.2	0.4	1.6	21.5
	cross-sectional	-24.1	-1.3	0.1	0.3	1.7	33.8
RegioPlus	time-series	-28.6	-1.8	0.2	1.6	2.4	210.2
	cross-sectional	-27.8	-2.2	0.3	2.1	2.8	236.4
region-branch	time-series	-94.5	-5.2	0.6	Inf	9.3	Inf
	cross-sectional	-92.8	-6.0	0.6	Inf	11.9	Inf
RegioPlus-branch	time-series	-94.5	-8.4	0.7	Inf	16.7	Inf
	cross-sectional	-94.0	-8.4	1.0	Inf	19.8	Inf

Table C.10 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

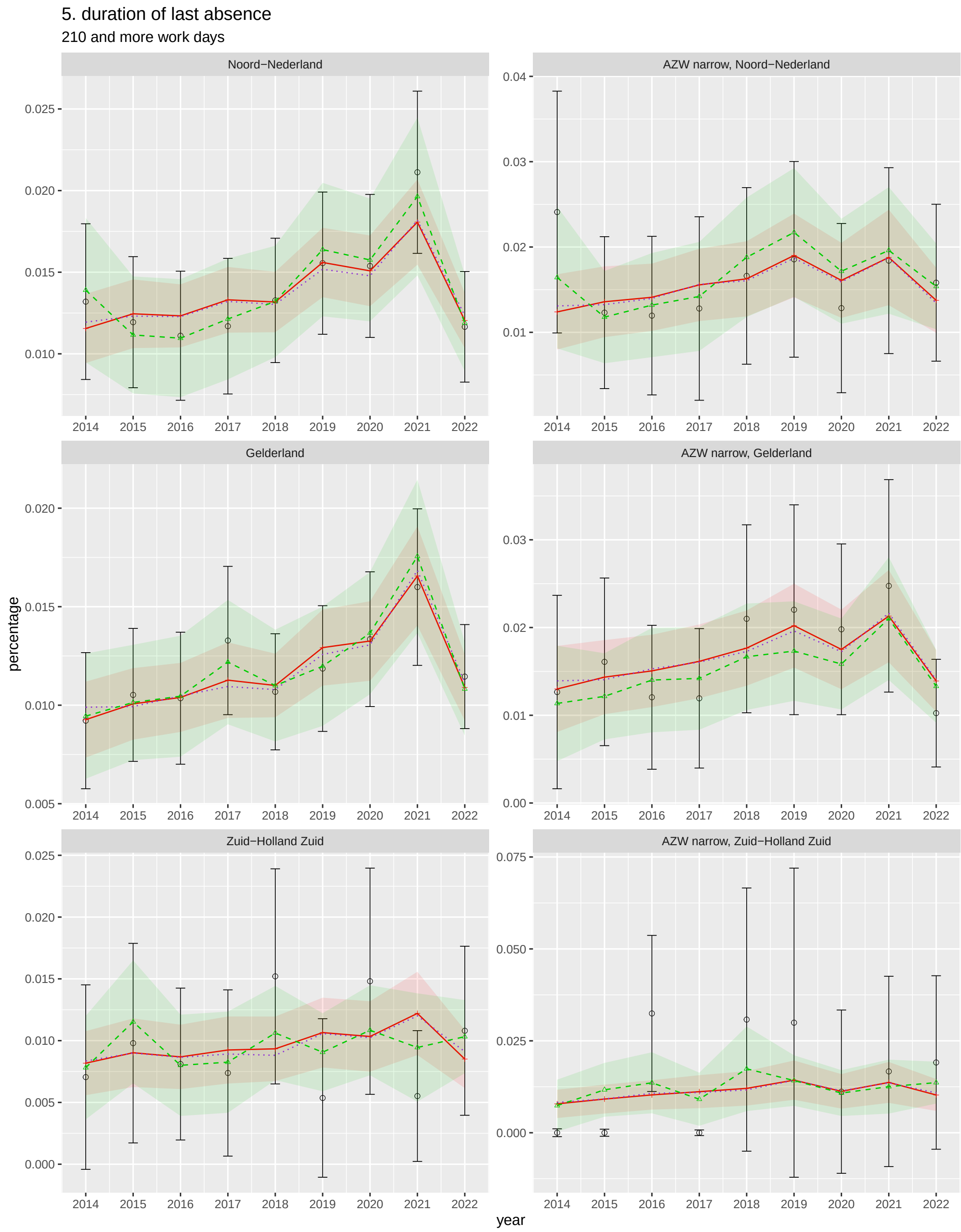

Figure C.16 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

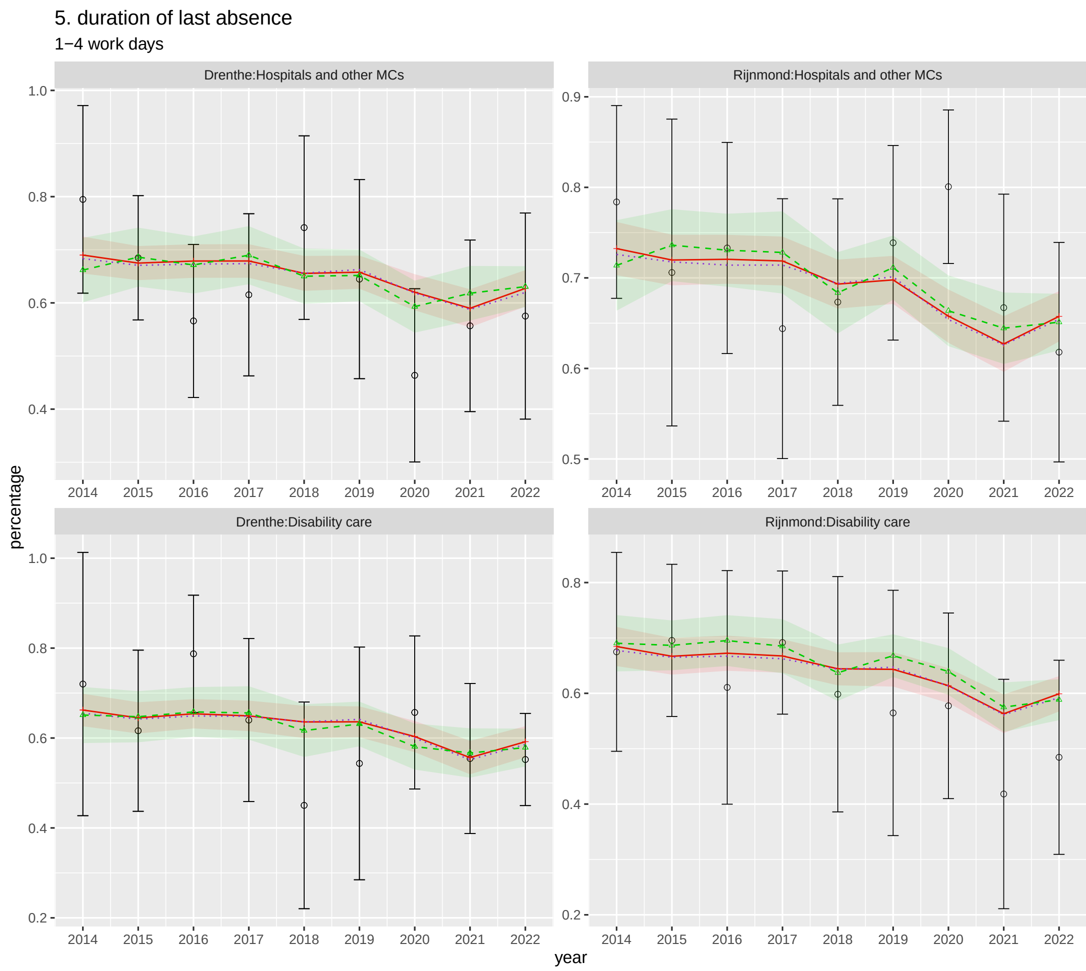

Figure C.17 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-93.6	-19.3	-1.5	14.2	32.9	299.3
	cross-sectional	-109.5	-23.7	-0.2	12.8	29.5	298.2
AZW branch	time-series	-94.1	-16.0	2.0	7.3	22.6	279.3
	cross-sectional	-132.4	-24.3	3.1	9.0	31.3	278.9
region	time-series	-72.6	-11.3	3.1	10.3	23.9	253.7
	cross-sectional	-72.5	-16.3	0.5	9.0	25.2	261.0
RegioPlus	time-series	-51.6	-13.3	2.4	6.5	21.4	166.0
	cross-sectional	-107.1	-18.2	3.2	6.5	25.3	176.0
region-branch	time-series	-87.3	-20.2	0.2	10.3	24.3	1274.5
	cross-sectional	-95.4	-22.3	0.6	12.4	27.4	1856.4
RegioPlus-branch	time-series	-263.3	-20.9	-1.7	19.0	22.5	1152.7
	cross-sectional	-265.4	-22.7	-1.3	22.1	25.2	1689.4

Table C.11 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-3.4	22.0	32.4	34.1	44.0	76.8
	cross-sectional	-5.2	3.1	9.1	13.2	22.3	54.8
AZW branch	time-series	-2.9	54.0	62.4	58.8	70.3	81.4
	cross-sectional	-8.1	34.6	48.1	44.9	58.4	76.6
region	time-series	16.1	45.8	55.1	54.1	62.4	77.1
	cross-sectional	0.0	10.4	34.7	28.2	39.8	56.6
RegioPlus	time-series	44.0	59.3	65.6	64.7	69.2	81.8
	cross-sectional	1.3	34.3	45.3	42.4	51.0	72.7
region-branch	time-series	-267.6	69.9	79.0	73.6	85.1	96.9
	cross-sectional	-636.1	57.3	69.9	60.8	78.7	95.4
RegioPlus-branch	time-series	-319.1	76.3	83.9	73.6	88.6	98.7
	cross-sectional	-607.9	65.1	77.2	61.4	84.8	98.3

Table C.12 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

C.6 Work-relatedness of the last absence

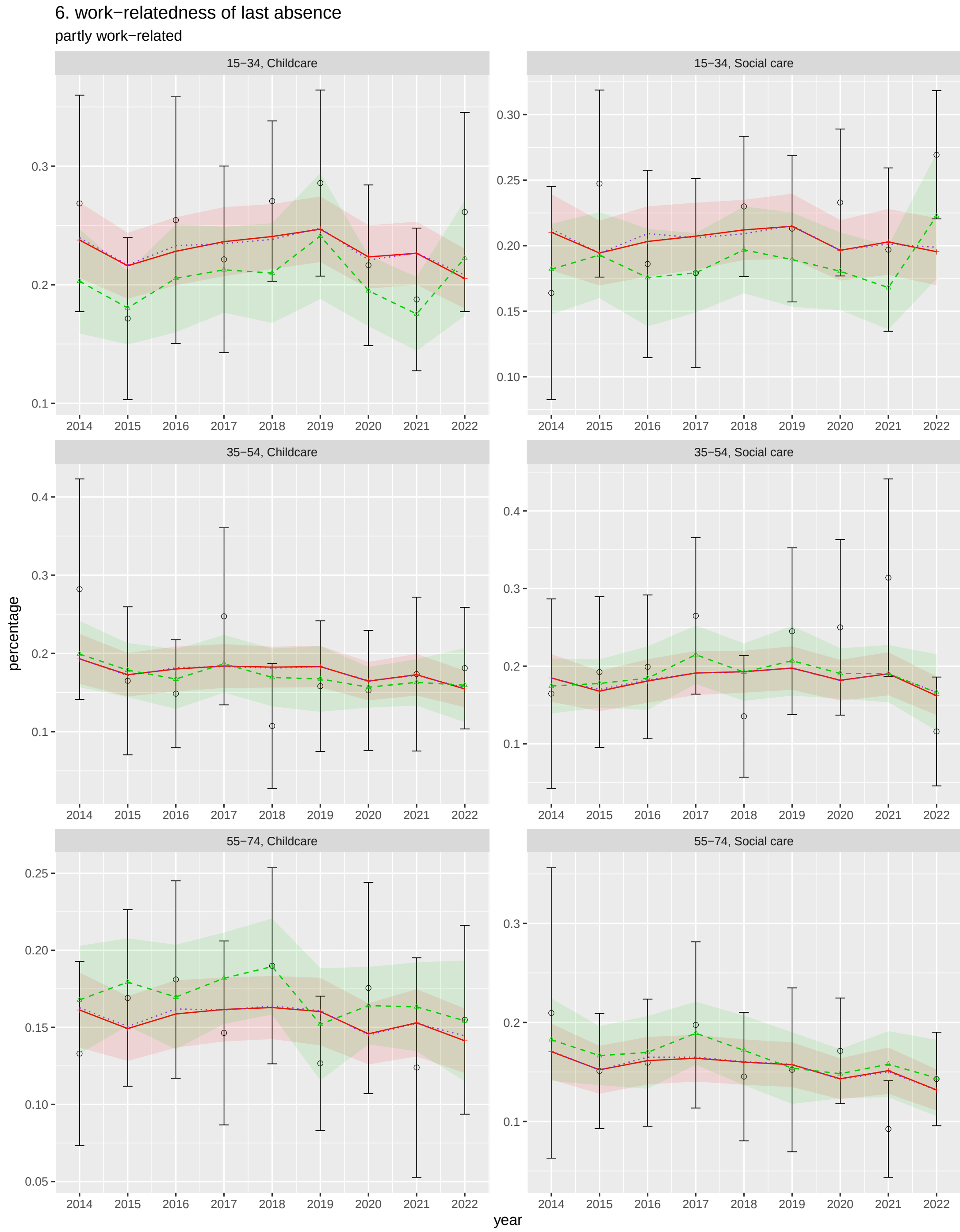

Figure C.18 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

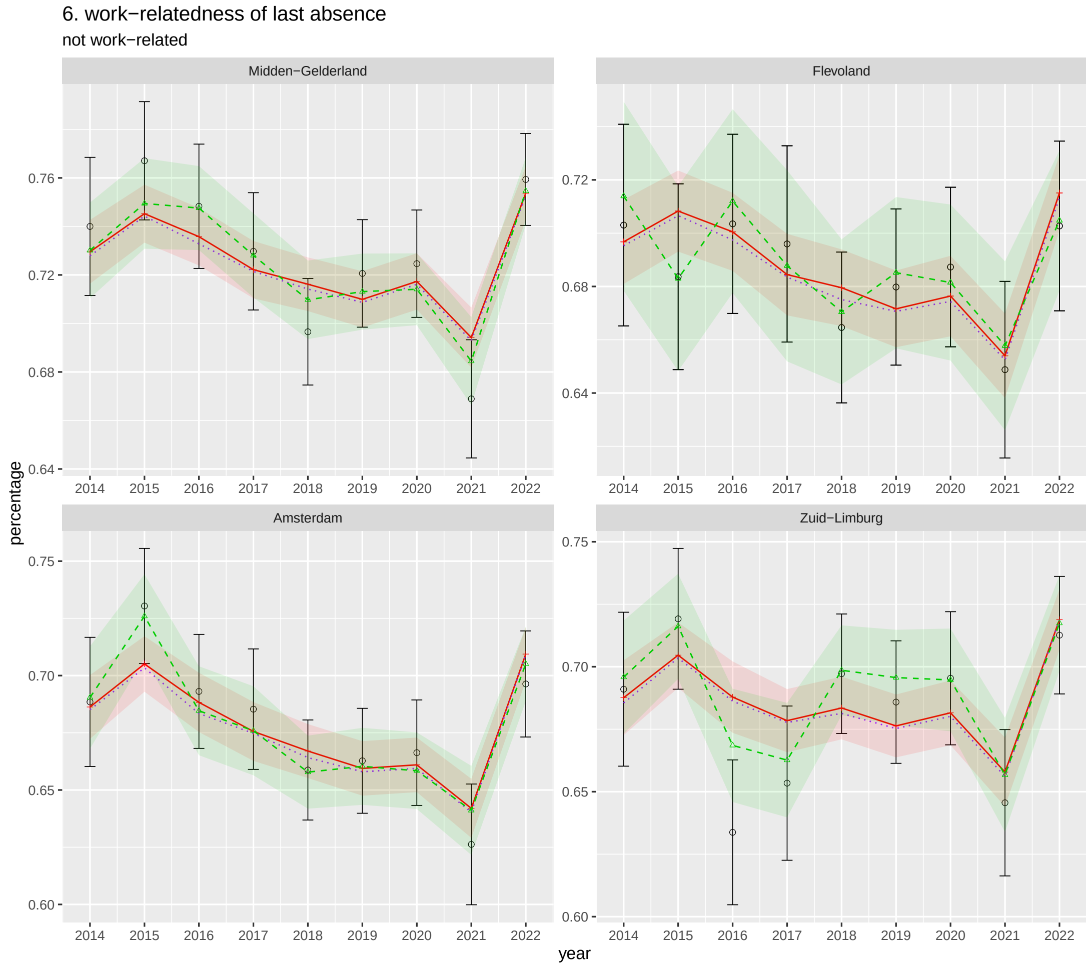

Figure C.19 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-9.1	-0.7	0.0	0.2	0.8	7.6
	cross-sectional	-9.7	-0.7	0.1	0.2	0.9	7.6
AZW branch	time-series	-29.3	-1.7	0.2	1.3	3.0	66.7
	cross-sectional	-24.9	-2.6	0.7	3.4	6.5	88.9
region	time-series	-11.3	-1.0	0.0	0.2	1.3	18.1
	cross-sectional	-10.3	-1.1	0.0	0.4	1.4	22.1
RegioPlus	time-series	-13.3	-1.8	0.1	0.5	1.9	25.3
	cross-sectional	-19.7	-2.1	0.2	0.7	2.7	25.2
region-branch	time-series	-76.2	-4.8	0.1	Inf	7.5	Inf
	cross-sectional	-74.3	-5.4	0.5	Inf	10.8	Inf
RegioPlus-branch	time-series	-83.2	-7.9	0.4	Inf	15.3	Inf
	cross-sectional	-81.6	-8.2	0.8	Inf	18.1	Inf

Table C.13 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

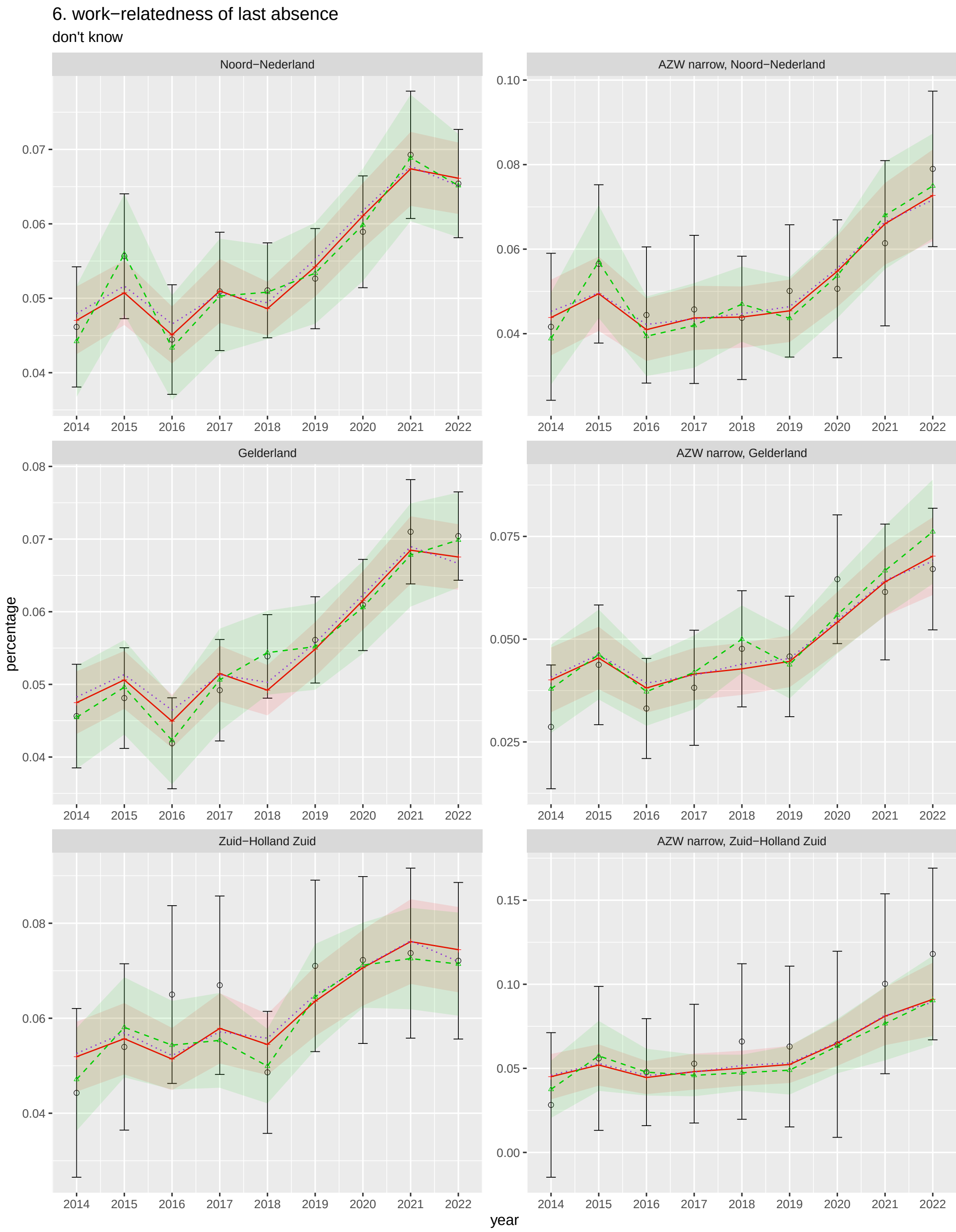

Figure C.20 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

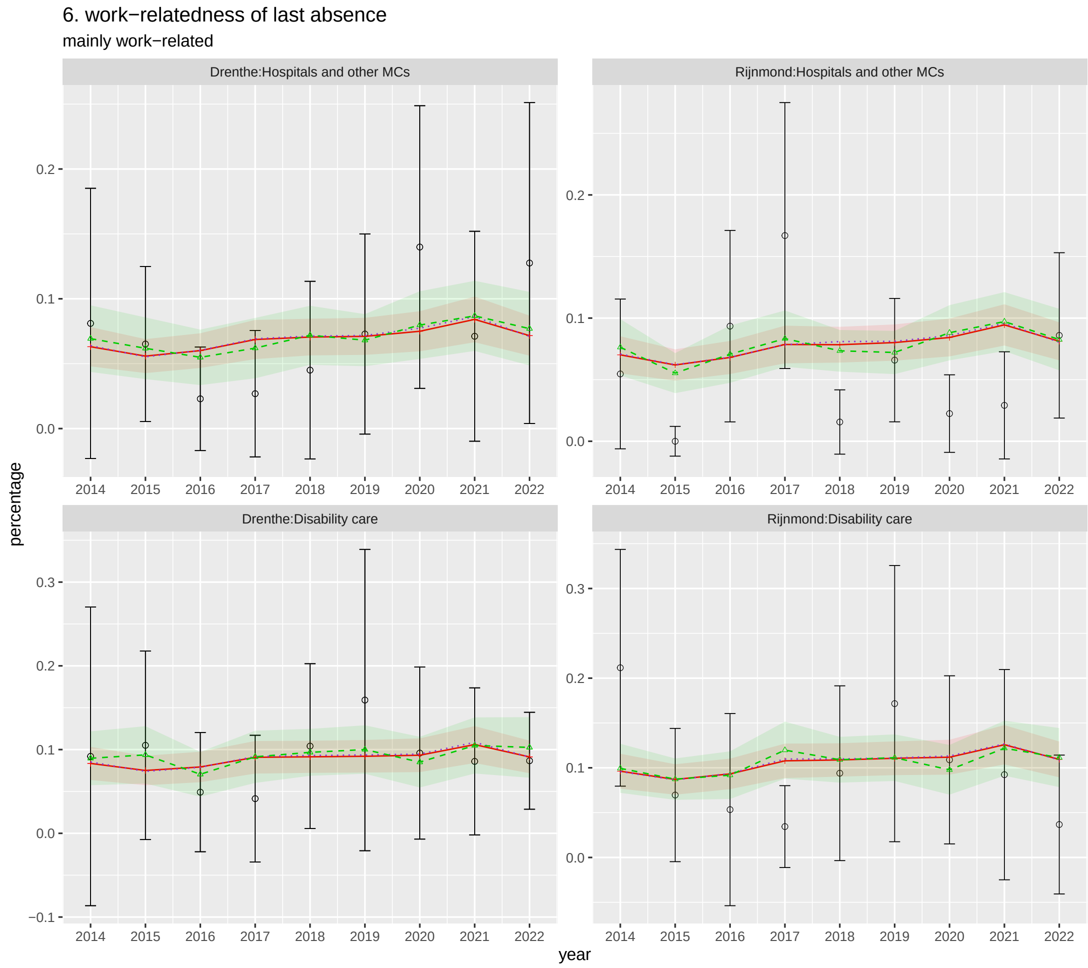

Figure C.21 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-76.6	-15.6	0.6	11.3	23.5	258.2
	cross-sectional	-91.7	-18.0	2.2	11.7	25.9	263.2
AZW branch	time-series	-89.7	-10.9	3.1	6.0	19.5	239.8
	cross-sectional	-138.2	-25.0	6.9	8.9	35.3	238.8
region	time-series	-55.8	-11.1	1.0	7.9	20.4	230.5
	cross-sectional	-62.5	-14.6	0.8	8.1	25.2	234.9
RegioPlus	time-series	-83.4	-15.9	1.1	4.7	22.1	140.5
	cross-sectional	-67.9	-19.3	3.6	5.1	25.1	153.4
region-branch	time-series	-126.4	-21.2	-1.5	2.5	20.8	243.7
	cross-sectional	-148.4	-24.2	-0.2	4.0	27.7	305.2
RegioPlus-branch	time-series	-283.0	-22.7	-2.9	2.0	19.5	731.0
	cross-sectional	-276.3	-23.3	-2.3	3.3	22.1	851.6

Table C.14 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-5.5	21.8	31.0	32.6	42.2	73.6
	cross-sectional	-11.3	-1.1	5.1	8.7	18.8	47.2
AZW branch	time-series	-5.2	52.8	62.3	57.8	68.9	82.4
	cross-sectional	-9.4	33.5	47.8	44.1	58.8	75.0
region	time-series	10.5	42.8	52.1	51.0	60.5	74.1
	cross-sectional	-6.2	3.8	31.8	24.6	38.1	53.3
RegioPlus	time-series	37.3	55.8	62.9	61.5	67.0	78.5
	cross-sectional	-0.3	31.9	45.4	41.2	50.5	72.6
region-branch	time-series	9.7	71.7	78.8	75.7	84.5	95.5
	cross-sectional	-4.3	59.8	69.9	65.2	78.6	93.5
RegioPlus-branch	time-series	-87.2	77.3	83.1	80.0	87.8	98.3
	cross-sectional	-107.4	68.8	77.5	72.4	84.3	97.9

Table C.15 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

C.7 Type of complaints of the last absence

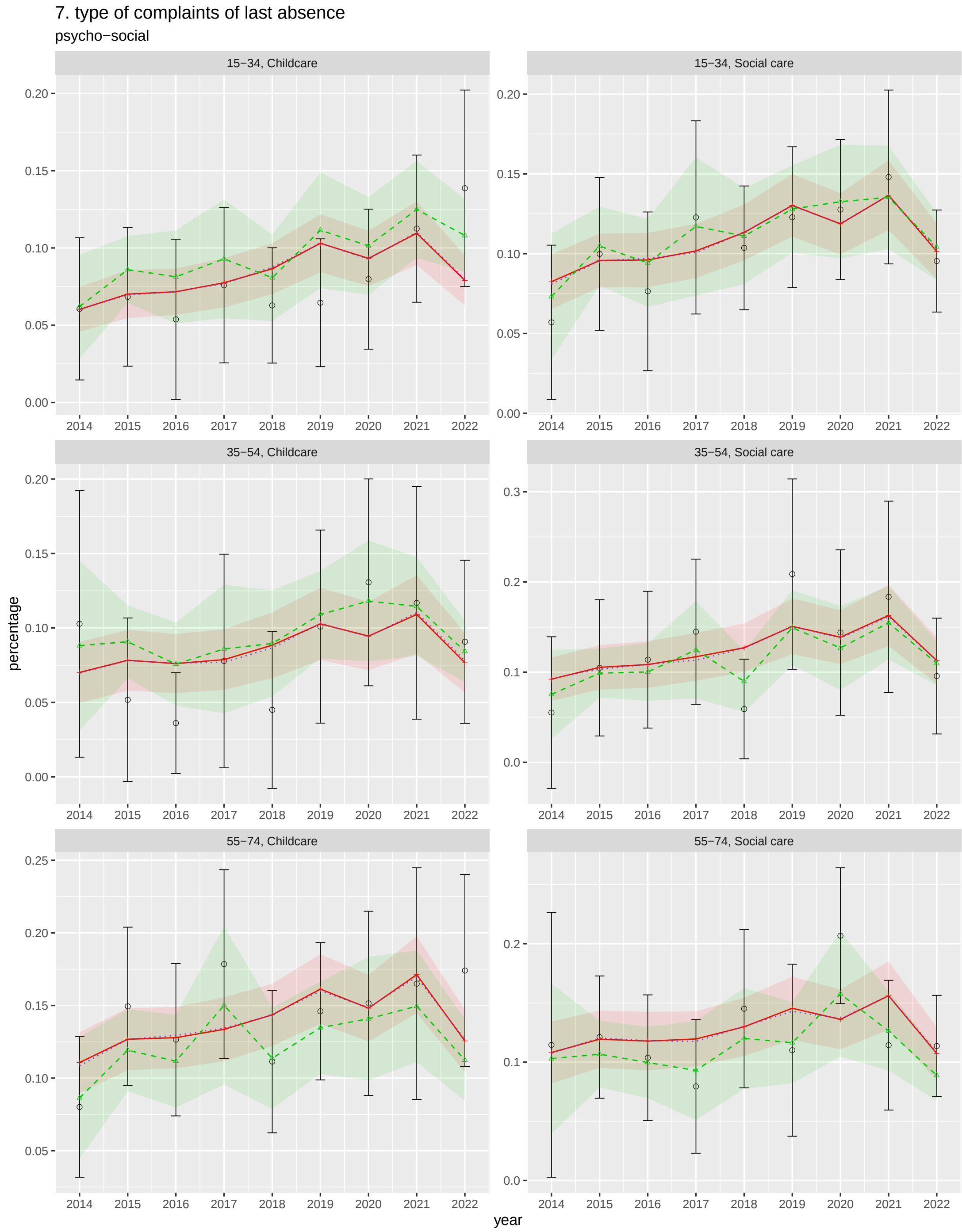

Figure C.22 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

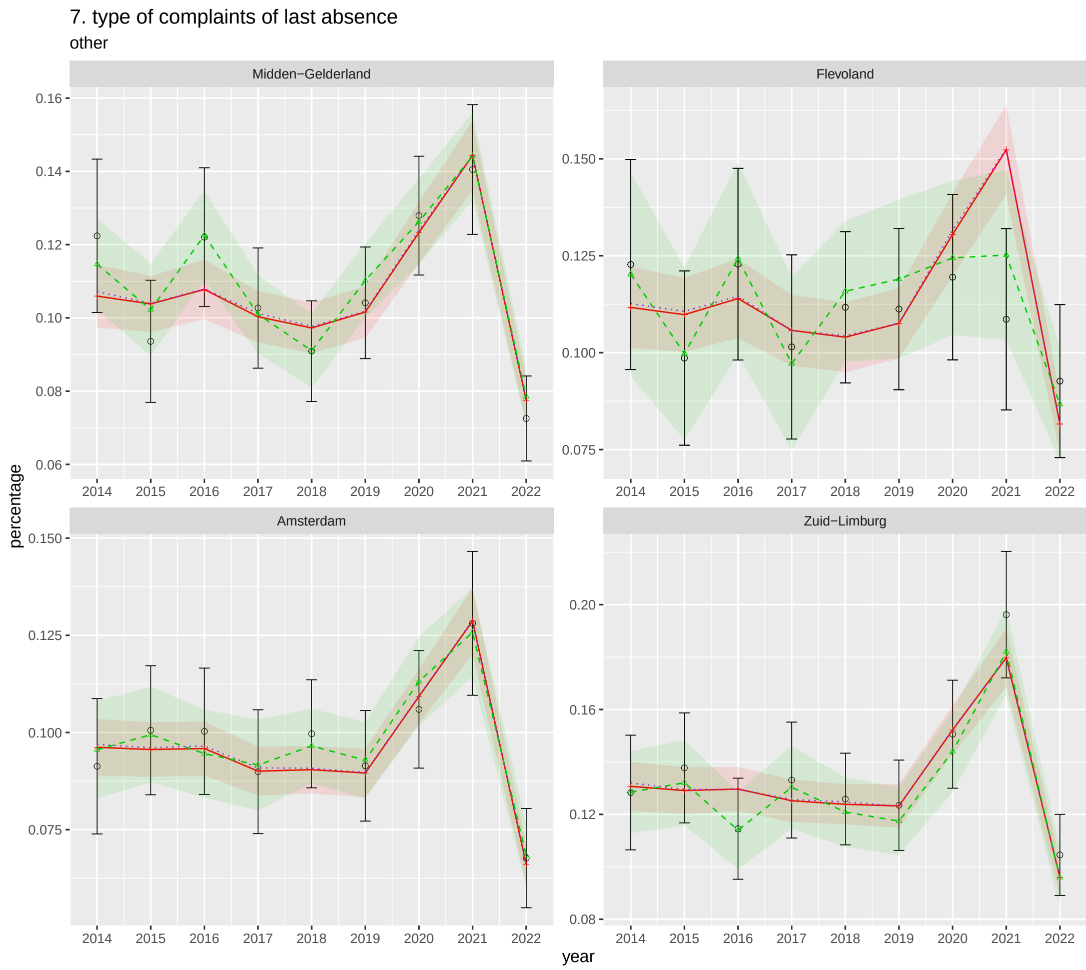

Figure C.23 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-9.9	-0.4	0.1	0.4	1.0	9.1
	cross-sectional	-9.6	-0.4	0.1	0.4	0.9	9.3
AZW branch	time-series	-21.3	-1.1	0.3	1.0	2.8	24.7
	cross-sectional	-38.1	-2.0	0.2	1.3	5.3	44.6
region	time-series	-5.6	-0.3	0.1	0.9	1.9	17.4
	cross-sectional	-6.8	-0.9	0.1	0.8	1.5	21.1
RegioPlus	time-series	-11.0	-0.8	0.3	0.9	2.4	17.4
	cross-sectional	-11.3	-1.0	0.3	0.8	2.5	21.1
region-branch	time-series	-60.2	-3.7	0.2	4.0	6.3	243.4
	cross-sectional	-58.4	-4.4	0.5	4.1	7.7	281.4
RegioPlus-branch	time-series	-76.7	-6.1	0.4	Inf	11.8	Inf
	cross-sectional	-74.2	-6.9	0.6	Inf	13.5	Inf

Table C.16 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

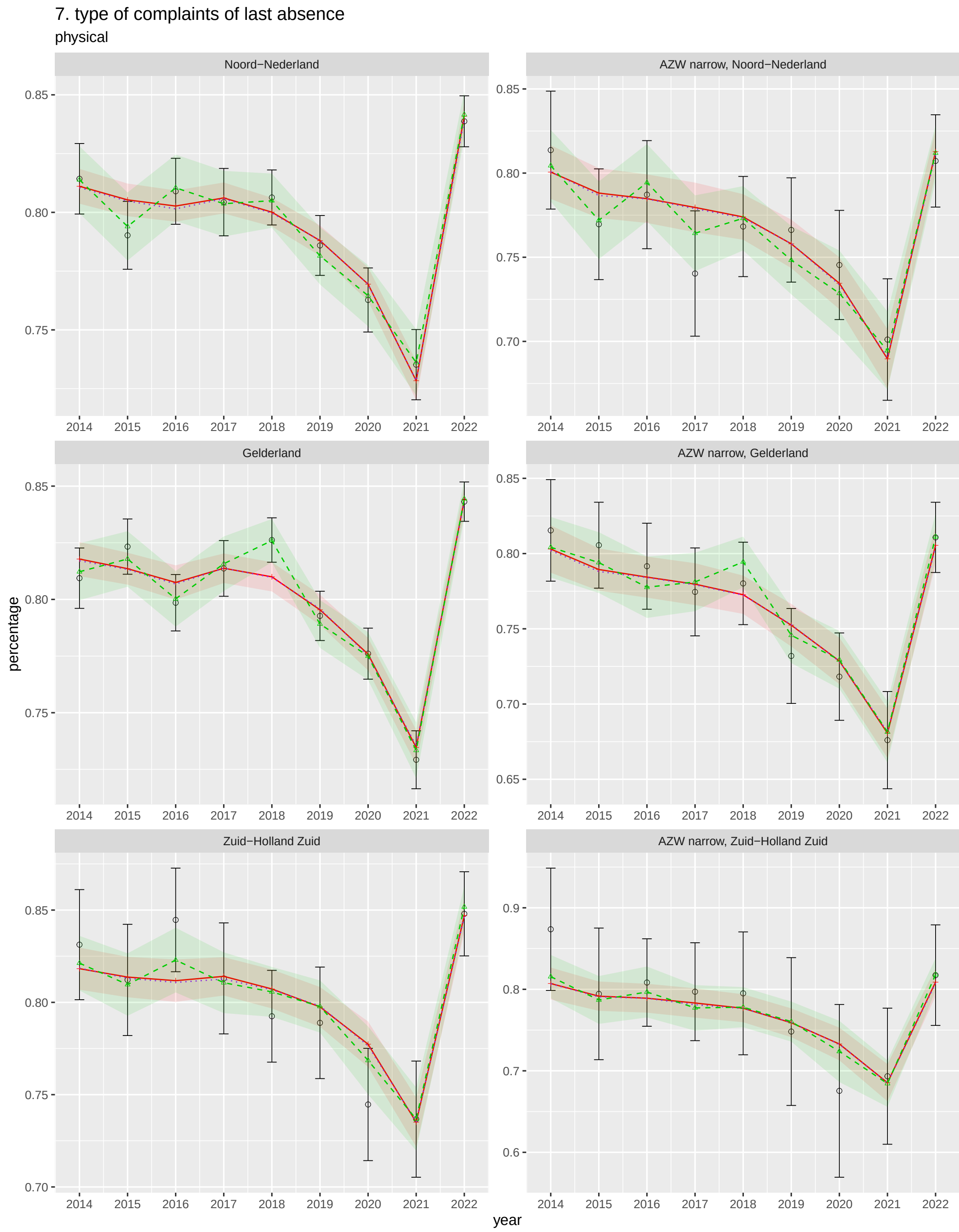

Figure C.24 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

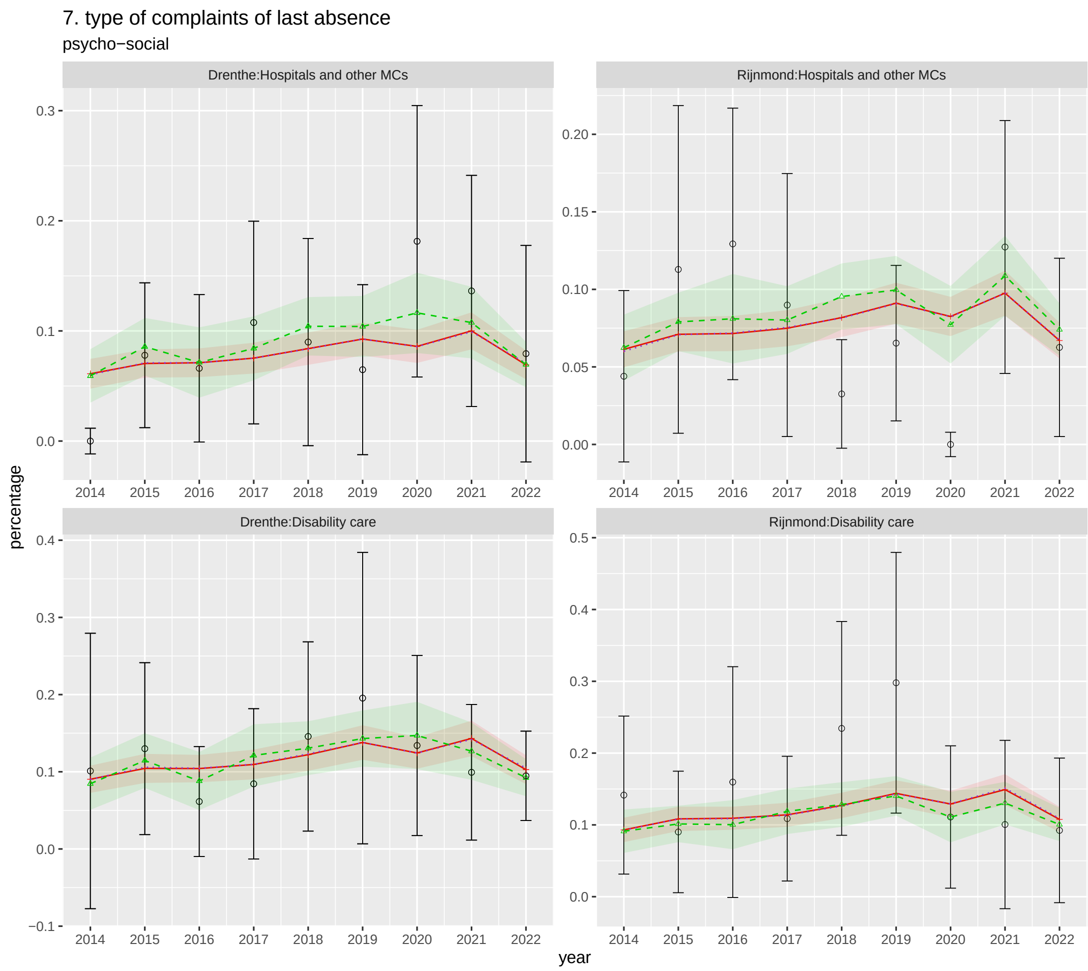

Figure C.25 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-69.6	-11.8	6.8	16.7	25.6	259.6
	cross-sectional	-72.4	-14.6	7.9	16.4	25.5	257.7
AZW branch	time-series	-68.9	-10.7	5.3	7.9	19.7	238.8
	cross-sectional	-133.0	-22.7	6.1	9.6	37.5	233.0
region	time-series	-54.6	-6.1	3.8	12.3	23.1	238.5
	cross-sectional	-49.9	-12.2	5.5	11.4	22.9	241.2
RegioPlus	time-series	-70.0	-11.3	5.4	7.0	21.3	142.8
	cross-sectional	-72.1	-15.2	5.3	6.8	23.4	156.7
region-branch	time-series	-160.1	-19.1	-0.2	2.1	20.4	225.5
	cross-sectional	-150.2	-22.9	0.4	2.7	23.1	210.1
RegioPlus-branch	time-series	-195.5	-22.6	-2.8	-1.1	17.3	593.4
	cross-sectional	-209.3	-23.8	-2.6	-0.6	20.2	659.7

Table C.17 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-4.0	22.9	33.7	34.8	44.9	75.5
	cross-sectional	-6.4	-0.4	4.9	9.1	18.8	48.4
AZW branch	time-series	-3.7	55.3	64.3	60.0	70.1	83.0
	cross-sectional	-6.0	33.3	47.7	43.7	57.8	76.0
region	time-series	15.8	43.8	53.6	53.1	62.3	75.4
	cross-sectional	-4.0	4.2	32.7	25.1	38.2	53.6
RegioPlus	time-series	49.3	58.7	66.4	65.0	69.9	80.1
	cross-sectional	-0.3	33.3	46.9	42.4	51.8	71.8
region-branch	time-series	14.6	74.4	81.1	77.8	85.9	94.8
	cross-sectional	-1.8	61.2	71.6	66.3	79.6	93.1
RegioPlus-branch	time-series	-49.9	81.1	85.5	83.1	89.2	96.9
	cross-sectional	-63.9	71.4	79.3	74.6	85.4	96.2

Table C.18 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

C.8 Most important reason that led to work-related complaints

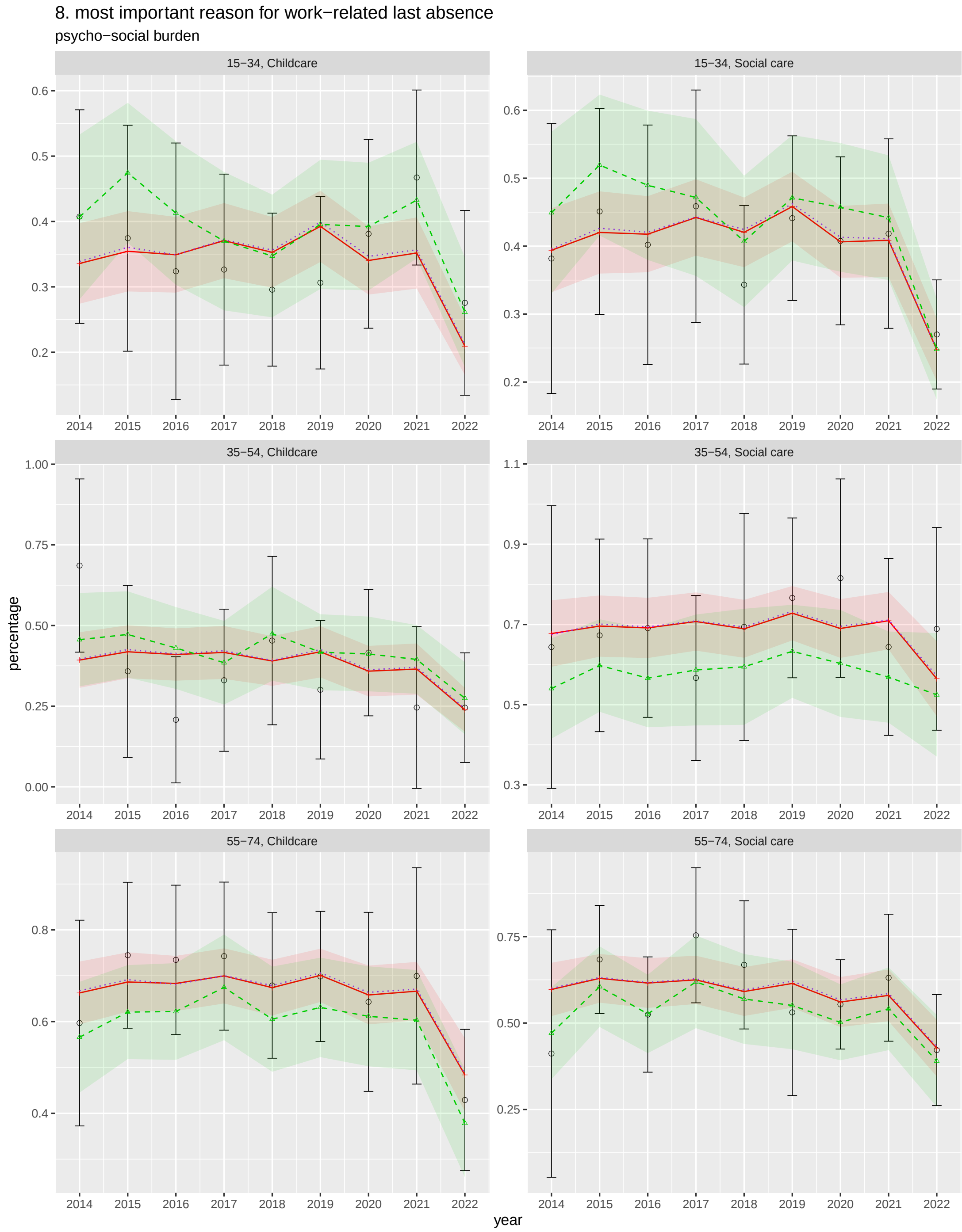

Figure C.26 Estimates and approximate 95% confidence intervals/bands, for a selection of two AZW branches by age class. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

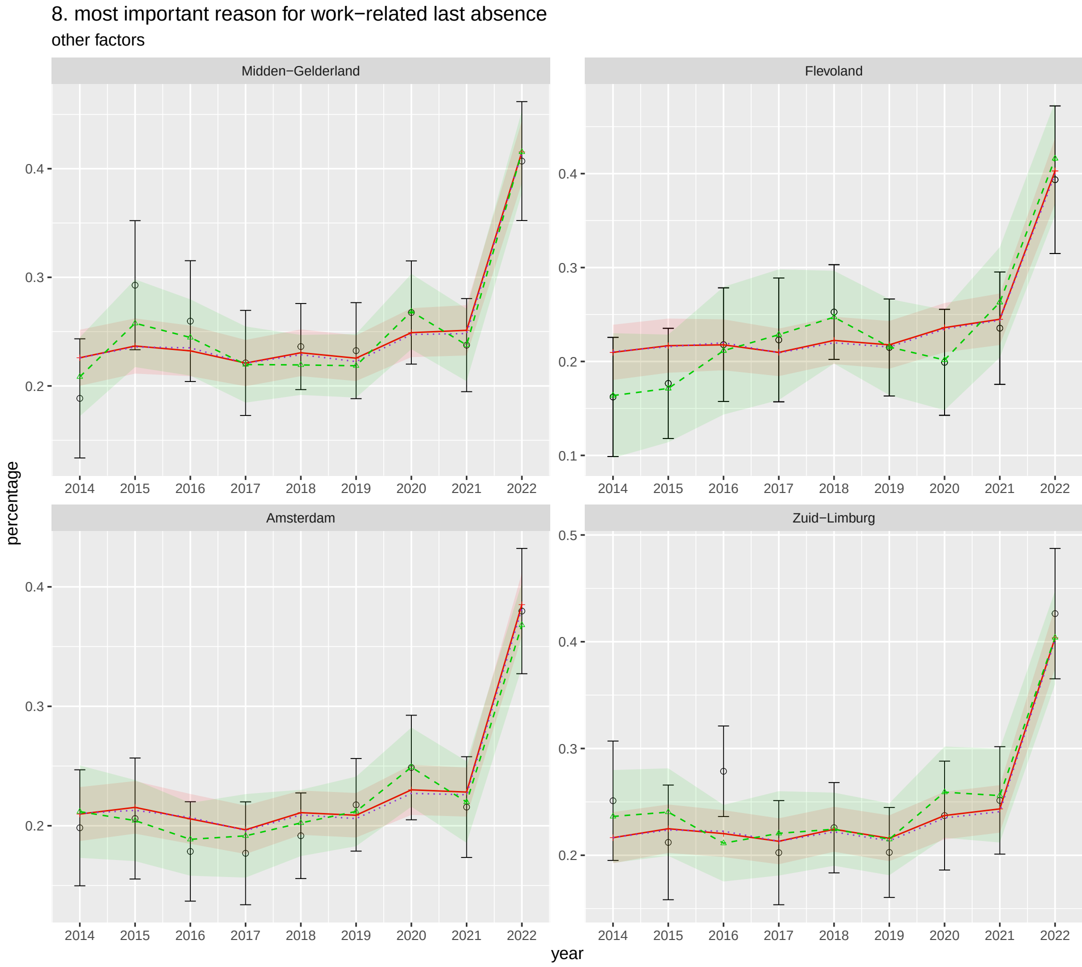

Figure C.27 Estimates and approximate 95% confidence intervals/bands, for a selection of 4 RegioPlus regions. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-2.4	-0.4	0.1	0.1	0.6	4.1
	cross-sectional	-3.8	-0.5	0.2	0.1	0.6	4.1
AZW branch	time-series	-24.5	-1.1	0.3	1.3	1.8	65.9
	cross-sectional	-27.0	-3.7	0.4	6.4	6.6	263.4
region	time-series	-13.9	-0.8	0.2	0.7	1.1	74.8
	cross-sectional	-15.8	-1.1	0.2	0.8	1.2	77.8
RegioPlus	time-series	-16.8	-1.5	0.2	0.5	1.8	74.8
	cross-sectional	-15.8	-1.7	0.1	0.6	1.7	77.8
region-branch	time-series	-68.4	-3.0	0.4	Inf	4.8	Inf
	cross-sectional	-59.9	-5.3	0.4	Inf	8.7	Inf
RegioPlus-branch	time-series	-74.6	-6.7	0.5	Inf	10.8	Inf
	cross-sectional	-68.3	-7.5	0.6	Inf	14.4	Inf

Table C.19 Mean relative bias (MRB) as a percentage of direct estimates averaged over time, by group of publication domains.

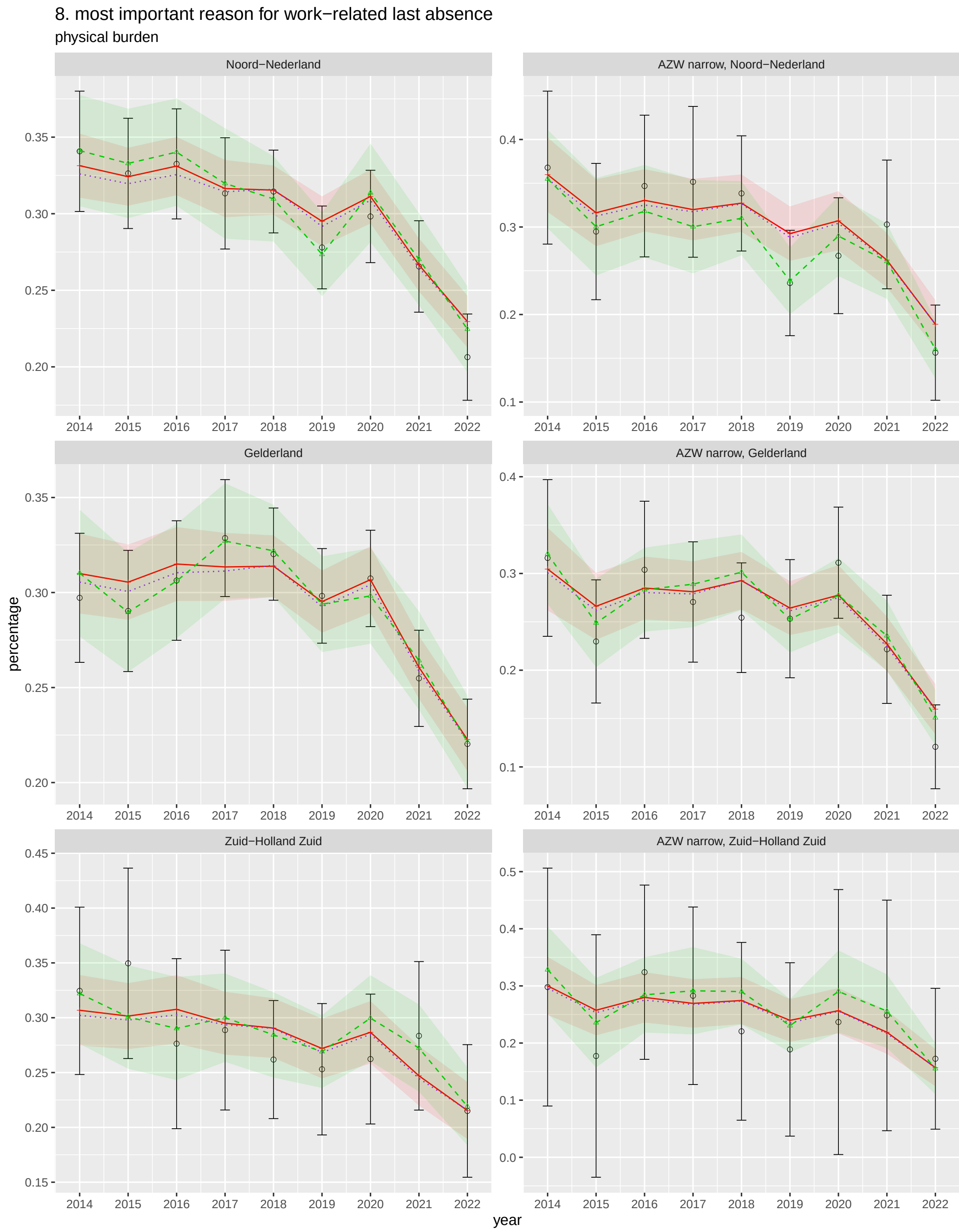

Figure C.28 Estimates and approximate 95% confidence intervals/bands, for a selection of regional domains both for all employees and the AZW narrow subpopulation. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

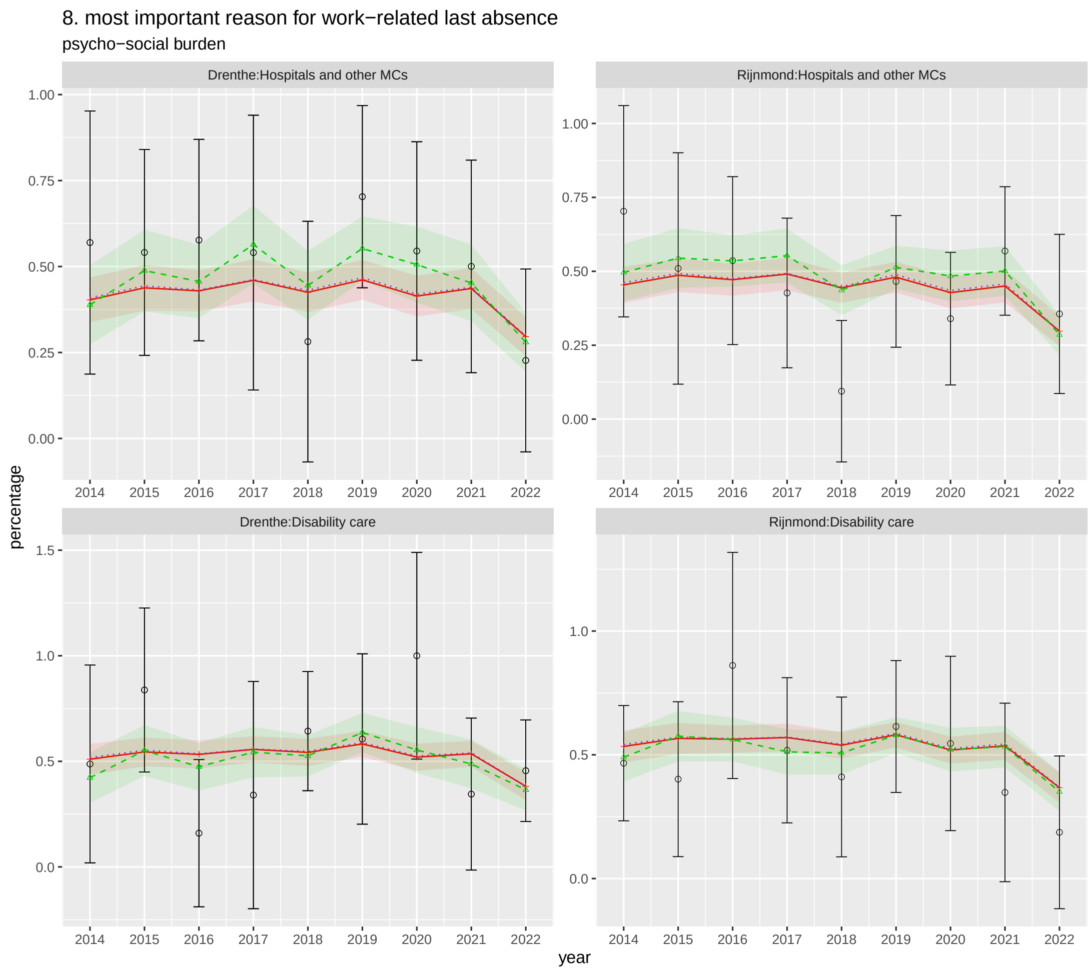

Figure C.29 Estimates and approximate 95% confidence intervals/bands, for a selection of domains at the most detailed output-level of RegioPlus and AZW branche combinations. In black: direct estimates, in red: time-series small area estimates, in green: cross-sectional small area estimates. The purple dotted line shows the benchmarked time-series small area estimates.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-71.4	-8.8	3.8	5.9	21.6	98.3
	cross-sectional	-75.8	-12.9	5.5	6.2	20.8	113.6
AZW branch	time-series	-89.3	-10.1	3.7	3.6	17.4	86.1
	cross-sectional	-216.8	-28.0	6.5	5.3	40.3	156.4
region	time-series	-96.4	-9.8	6.5	5.8	18.5	138.4
	cross-sectional	-93.1	-19.5	3.5	4.6	25.4	144.0
RegioPlus	time-series	-91.4	-13.5	3.4	3.2	19.2	138.4
	cross-sectional	-92.7	-15.9	2.4	2.9	22.4	144.0
region-branch	time-series	-157.6	-13.6	4.2	5.3	24.0	191.4
	cross-sectional	-145.3	-18.6	6.5	6.0	33.1	169.1
RegioPlus-branch	time-series	-369.7	-14.8	5.3	7.7	28.2	542.7
	cross-sectional	-328.6	-17.1	5.9	8.3	32.6	445.4

Table C.20 Mean relative bias (MRBse) as a percentage of the standard error of the time-averaged direct estimate, by group of publication domains.

group	model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
total	time-series	-4.1	23.8	32.3	34.0	42.3	72.3
	cross-sectional	-10.9	-1.5	3.2	7.4	17.0	43.4
AZW branch	time-series	-1.5	55.8	63.6	59.5	70.2	84.2
	cross-sectional	-8.0	21.0	37.0	34.9	48.3	76.5
region	time-series	14.9	42.6	52.6	51.5	61.4	74.0
	cross-sectional	-4.8	3.8	31.5	23.6	36.6	53.0
RegioPlus	time-series	41.9	55.8	64.0	62.5	68.0	80.1
	cross-sectional	-0.4	31.0	44.0	40.6	50.3	71.8
region-branch	time-series	14.8	73.7	80.9	77.2	86.1	95.7
	cross-sectional	-0.7	57.2	68.5	63.5	78.1	94.5
RegioPlus-branch	time-series	13.3	80.0	85.3	81.9	89.0	95.7
	cross-sectional	6.7	69.5	77.8	72.7	83.6	94.5

Table C.21 Mean relative reduction of direct standard error (RRSE) averaged over time, by group of publication domains.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information