A flexible framework for interpretable and individualized reporting of model results

Hannah Schulz-Kümpel, Sabine Hoffmann

September 5th, 2023

CEN2023 | Young Statisticians Session

Motivation

- One central purpose of (biomedical) research: Facilitate evidence based decision making.
- This requires not only appropriate study designs but also practically usable results.

Practically usable results $\hat{=}$ numerical quantities from which specific decision rules may be derived, i.e. that are *interpretable* and *individualizable*

- Relevance is rising especially in the context of personalized medicine.

Reporting methods are adapting

Conventional:

One or two standard quantities that are always reported.

Problematic examples:

- Fixation on p-values
- OR being commonly interpreted as RR

More recently:

More interpretable quantities like ITEs and change in probability are being proposed.

However:

- Often heuristically or narrowly defined
- Usually only one correct interpretation per quantity

Making the solution fit the problem:

Interpretable effect measures that may flexibly be adjusted to answer specific questions are needed.

The framework's setting I

- The proposed framework applies to any (thus far parametric, i.e. finite dimensional parameter space Θ) models where the conditional expectation of the target variable can be written as
$$\mathbb{E}[Y|X] = g_\theta(X)$$
where the function $g_\theta : \mathbb{R}^p \longrightarrow \mathbb{R}$, $p \in \mathbb{N}$, is at least once partially differentiable w.r.t. each metric element of X $\forall \theta \in \Theta$.

- Here, we consider the function g_θ to be indexed by $\theta \in \Theta$ in the sense of the following mapping
$$g : \Theta \longrightarrow \mathcal{M}(\mathbb{R}^p), \quad \theta \longmapsto g_\theta$$
with $\mathcal{M}(\mathbb{R}^p) := \left\{f : \mathbb{R}^p \longrightarrow \mathbb{R} | \ f \text{ is } \big(\mathcal{B}(\mathbb{R}^p), \mathcal{B}(\mathbb{R})\big)\text{-measurable}\right\}$, $p \in \mathbb{N}$.

The framework's setting II

- Given this, we propose to utilize probability measures to derive appropriately weighted means of functions of g_θ over areas of interest.

- Specifically, for some function $h : \mathbb{R}^d \longrightarrow \mathbb{R}$, $d \in \mathbb{N}_{>0}$, and set $D \subseteq \mathbb{R}^{\tilde{d}}$, $d \geq \tilde{d} \in \mathbb{N}_{>0}$, we consider probability measures μ that satisfy the following requirements:

(M1) μ is a probability measure on $\left(\mathbb{R}^{\tilde{d}}, \mathcal{B}(\mathbb{R}^{\tilde{d}})\right)$.

(M2) $\mathrm{supp}(\mu) \subseteq D$, if required as a result of μ being normalized w.r.t. D.

(M3) $\int_{\mathbb{R}^{\tilde{d}}} |h(x)| d\mu(x) < \infty$, if $\tilde{d} = d$, or $\forall x_a \in \mathbb{R}^{d-\tilde{d}} : \int_{\mathbb{R}^{\tilde{d}}} |h(x_a, x_b)| d\mu(x_b) < \infty$, if $\tilde{d} < d$.

Quantities in the framework

- **generalized marginal effects** - change of conditional expectation, averaged over a certain set of regressor/feature values.
- **individualized expectation** - conditional expectation, averaged over a certain set of regressor/feature values.
- **individualized predictive distribution** - distribution of the target variable with a specific individualized expectation set as expected value.
  $\longrightarrow$ combines estimation and sampling uncertainty.

Advantages compared to existing effect size estimates and visualizations

1. For each definition, the user can individually determine probability measures that represent the situationally appropriate averaging over the inputs.

$\longrightarrow$ Quantities may be seen as *tool kits* that can be specified to derive effects/expectations etc. over areas of interest.

2. After specification, one gets functions of the parameter vector θ for each quantity.

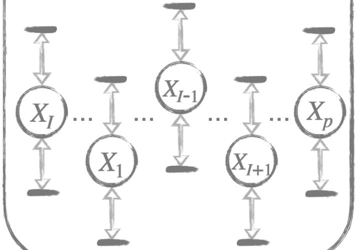
$\longrightarrow$ This allows us to provide a consistent method of calculating point estimates and well-interpretable uncertainty regions.

Assumptions

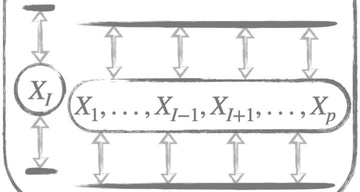
In our framework, a regressor/feature whose effect on the target variable is of interest (X_{I}) has to be chosen for each quantity.

To aide in the choices of probability measures, we have identified the following "assumptions":

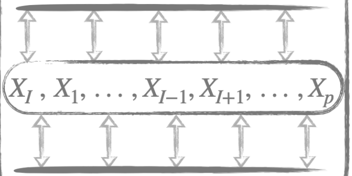
Assumption (A.I) :
All regressors are independent variables.



Assumption (A.II') :
While the regressors follow a joint distribution, that of interest is treated as independent.



Assumption (A.II'') :
The regressors follow a joint distribution and are treated as such.



Choices of probability measures

One of the main contributions of the proposed framework is its flexibility to be tailored to most research settings, so there are no universally best choices of probability measure.

Still, here are some examples that will often be reasonably applicable:

- The distribution given by the **relative frequency** in a given data set.
- The (discrete) **Uniform distribution** - for metric regressors this choice additionally has nice computational properties.
- The distribution of characteristics in the population as **determined by previous studies**.

$\Rightarrow$ Each combination of assumption and marginal probability measure choice leads to a different interpretation of the resulting quantity!

Uncertainty modelling

- We take a Bayesian approach to quantifying uncertainty – but all methods can still be applied to the results of both Bayesian and frequentist analyses!

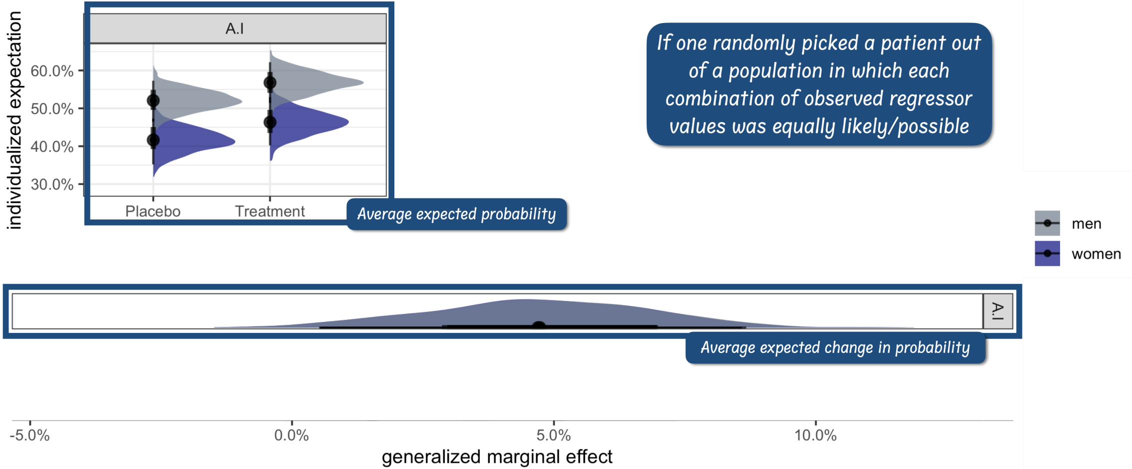
- Specifically, this is achieved by treating θ as a random variable with distribution equal to
  - the posterior distribution given by a Bayesian analysis
  - $N(\hat{\theta}, \Sigma_{\hat{\theta}})$ for the frequentist point estimate $\hat{\theta}$ and covariance matrix $\Sigma_{\hat{\theta}}$

  and then deriving a point estimate (e.g. *median*) and credible set (e.g. *equal-tailed interval*) directly for the random variable defined by $q \circ \theta$, for any considered quantity q.

Example: International Stroke Trial; 30-60 year old patients

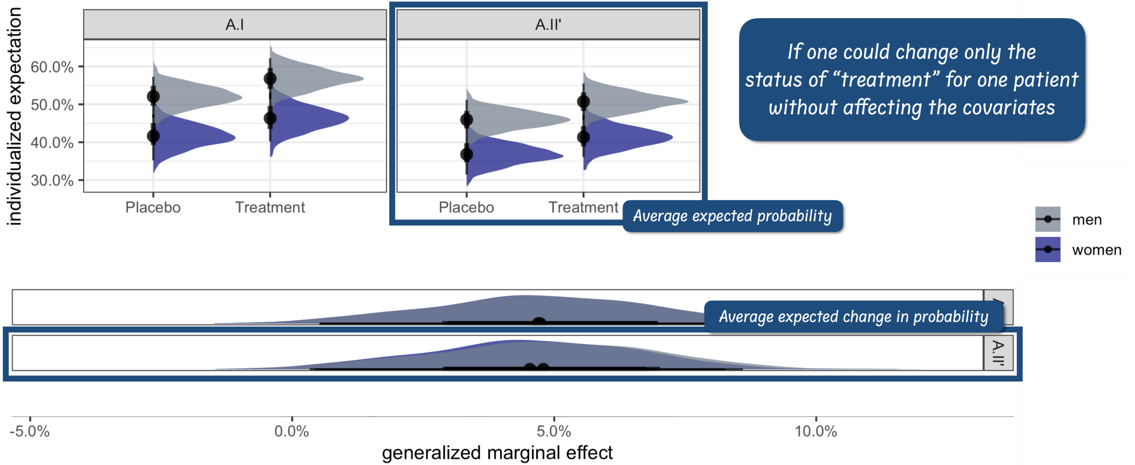
Under assumption (A.I), with the probability measure corresponding to the relative frequency:

If one randomly picked a patient out of a population in which each combination of observed regressor values was equally likely/possible



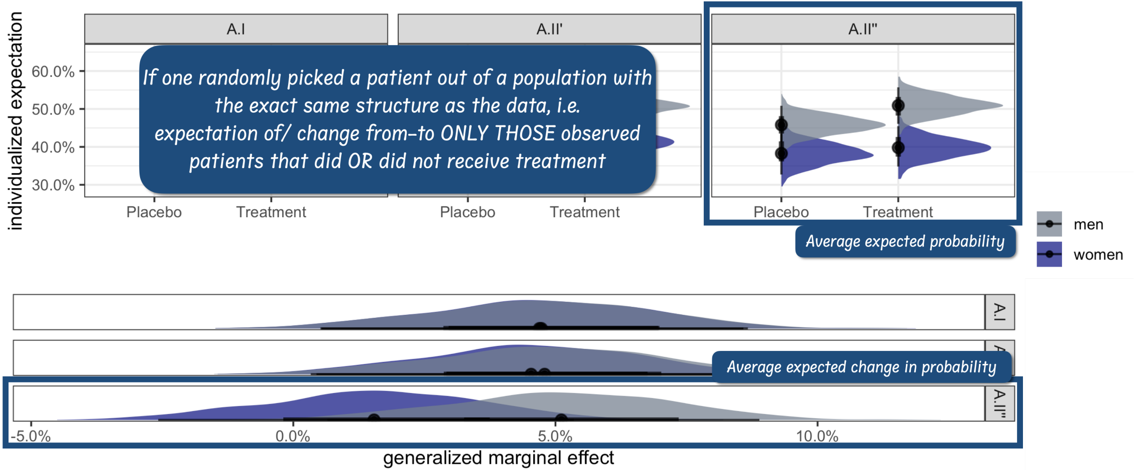
Example: International Stroke Trial; 30-60 year old patients

Under assumption (A.II'), with the probability measure corresponding to the relative frequency:



Example: International Stroke Trial; 30-60 year old patients

Under assumption (A.II"), with the probability measure corresponding to the relative frequency:



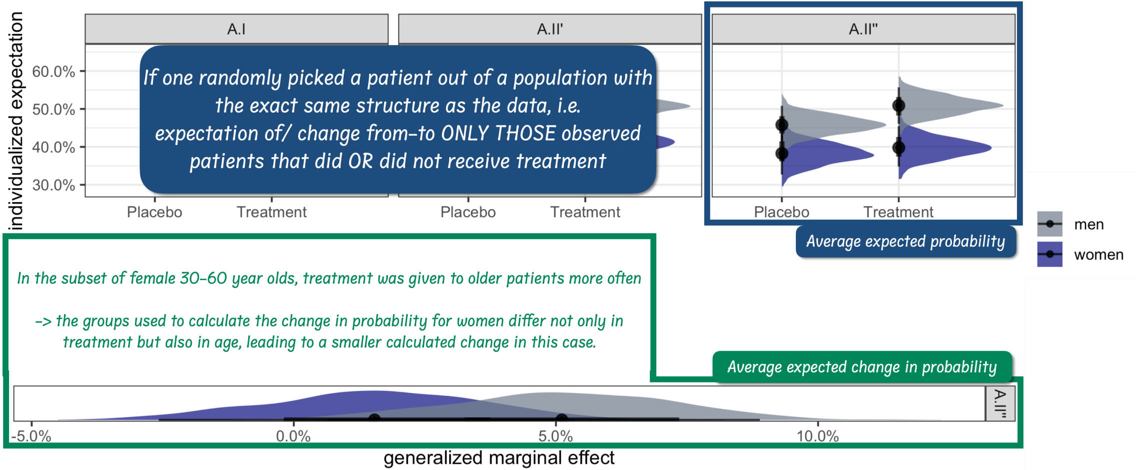
If one randomly picked a patient out of a population with the exact same structure as the data, i.e. expectation of/ change from-to ONLY THOSE observed patients that did OR did not receive treatment

Average expected probability

Average expected change in probability

Example: International Stroke Trial; 30-60 year old patients

Under assumption (A.II”), with the probability measure corresponding to the relative frequency:



If one randomly picked a patient out of a population with the exact same structure as the data, i.e. expectation of/ change from-to ONLY THOSE observed patients that did OR did not receive treatment

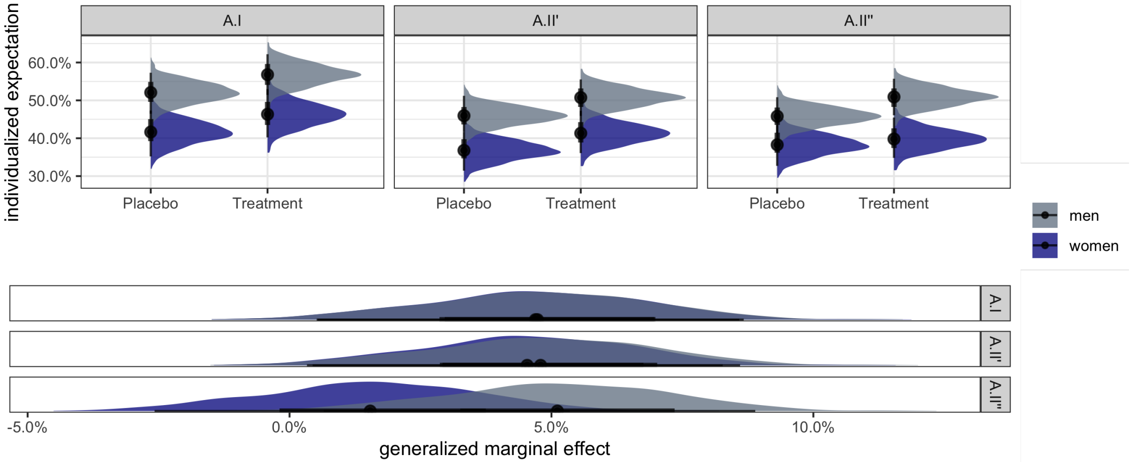
Average expected probability

In the subset of female 30-60 year olds, treatment was given to older patients more often

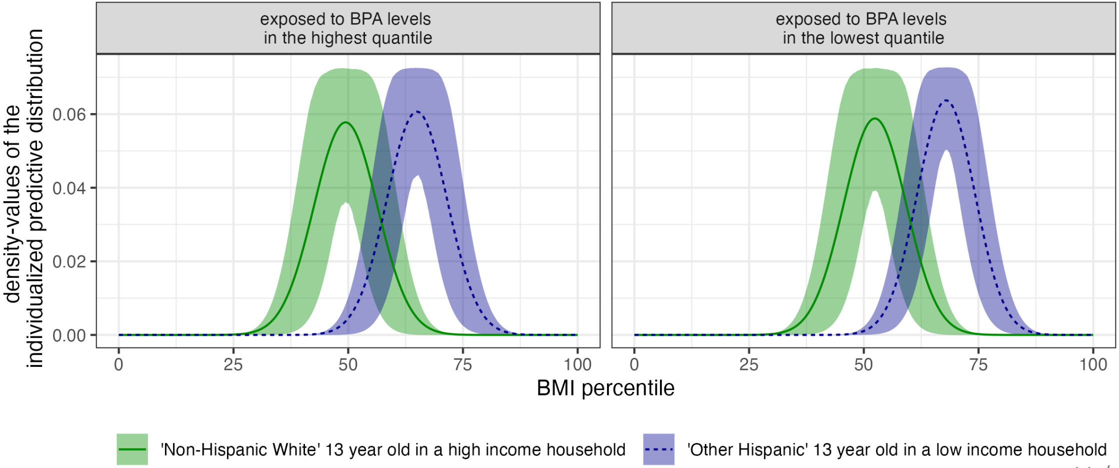
-> the groups used to calculate the change in probability for women differ not only in treatment but also in age, leading to a smaller calculated change in this case.

Average expected change in probability

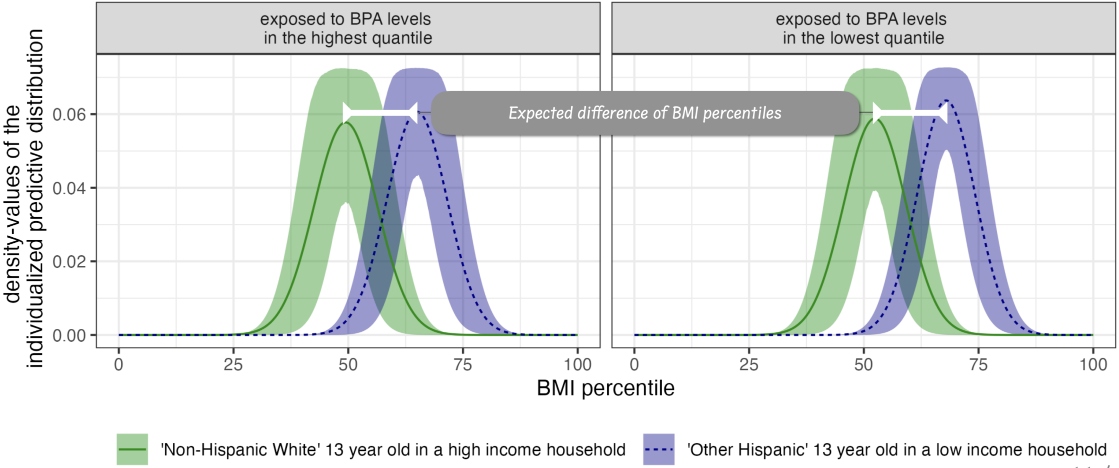
Example: International Stroke Trial; 30-60 year old patients



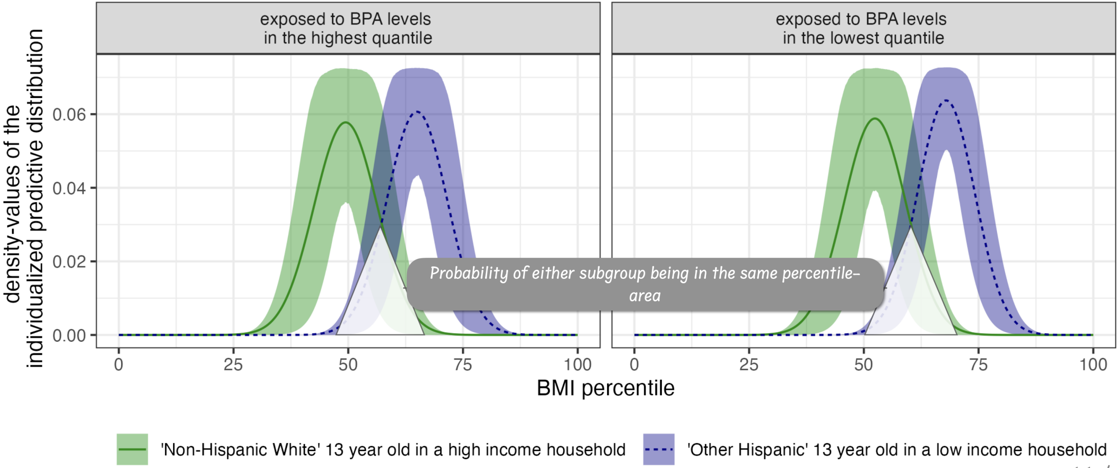
Example: Association of exposure to BPA and obesity in children & adolescents (NHANES 2005-2010)



Example: Association of exposure to BPA and obesity in children & adolescents (NHANES 2005-2010)



Example: Association of exposure to BPA and obesity in children & adolescents (NHANES 2005-2010)



Outlook

- We are currently working on extending the framework to non- and semi-parametric settings.

- Another natural extension of the given framework is to quantify a broader concept of change, both *absolute* and *relative*, as a random variable of the form the random variable

$$Y^{[\text{change}]} := c\left(Y^{[\text{new}]}, Y^{[\text{base}]}\right),$$

$$\text{with} \quad Y^{[\text{base}]} \sim p_{\mu=indiv.\,exp.\,1}(y|\theta) \quad \text{and} \quad Y^{[\text{new}]} \sim p_{\mu=indiv.\,exp.\,2}(y|\theta),$$

with $c : \mathbb{R} \longrightarrow \mathbb{R}$ representing some change function, such as the *difference* or *quotient*.

References

Gelman, Andrew and Iain Pardoe (2007). "2. Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components". In: *Sociological Methodology* 37.1, pp. 23–51. DOI: 10.1111/j.1467-9531.2007.00181.x.

Goldstein, Alex et al. (2015). "Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation". In: *Journal of Computational and Graphical Statistics* 24.1, pp. 44–65. DOI: 10.1080/10618600.2014.907095.

Kümpel and Hoffmann (2022). "A formal framework for generalized reporting methods in parametric settings". In: arXiv: 2211.02621 [stat.ME].

Sandercock, P et al. (2016). "The third International Stroke Trial (IST-3), 2000-2015 [dataset]". In: *University of Edinburgh & Edinburgh Clinical Trials Unit*. URL: https://doi.org/10.7488/ds/1350.

Silberzahn, R. et al. (2018). "Many analysts, one data set: Making transparent how variations in analytic choices affect results". In: *Advances in Methods and Practices in Psychological Science* 1, pp. 337–356. DOI: https://doi.org/10.1177/2515245917747646.

Williams, R. (2012). "Using the margins command to estimate and interpret adjusted predictions and marginal effects". In: *Stata Journal* 12.2, 308–331(24).

Wu, Bangsheng et al. (Aug. 2020). "Using three statistical methods to analyze the association between exposure to 9 compounds and obesity in children and adolescents: NHANES 2005-2010.". In: *Environ Health* 19.1, p. 94.

Deutsche Region der
Internationalen Biometrischen Gesellschaft
(IBS-DR)

LMU Open Science Center

Workshop on
"Open Replicable Research"
5th and 6th of October in Munich

Topics: Data sharing and Replicability

Invited Speakers: Ulrich Dirnagl (QUEST Berlin)
Ioana Cristea (Padova)

Abstract submission until the 10th of September